# THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE
OF MATHEMATICAL STATISTICS

## *Contents*

Vol. XVIII, No. 2 — June, 1947

# THE ANNALS OF MATHEMATICAL STATISTICS

The Annals of Mathematical Statistics is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore 2, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the Annals of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore 2, Md., or to the Secretary of the Institute of Mathematical Statistics, P. S. Dwyer, 116 Rackham Hall, University of Michigan, Ann Arbor, Mich.

Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December.

Manuscripts for publication in the Annals of Mathematical Statistics should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the Annals is $5.00 per year. Single copies $1.50. Back numbers are available at $5.00 per volume, or $1.50 per single issue.

Composed and Printed at the
WAVERLY PRESS, Inc.
Baltimore, Md., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879

# PROBLEMS IN PROBABILITY THEORY

By Harald Cramér

*University of Stockholm*

**1. Introduction.** The following survey of problems in probability theory has been written for the occasion of the Princeton Bicentennial Conference on "The Problems of Mathematics," Dec. 17–19, 1946. It is strictly confined to the purely mathematical aspects of the subject. Thus all questions concerned with the philosophical foundations of mathematical probability, or with its ever increasing fields of application, will be entirely left out.

No attempt to completeness has been made, and the choice of the problems considered is, of course, highly subjective. It is also necessary to point out explicitly that the literature of the war years has only recently—and still far from completely—been available in Sweden. Owing to this fact, it is almost unavoidable that this paper will be found incomplete in many respects.

## I. FUNDAMENTAL NOTIONS

**2. Probability distributions.** From a purely mathematical point of view, probability theory may be regarded as the theory of certain classes of *additive set functions*, defined on spaces of more or less general types. The basic structure of the theory has been set out in a clear and concise way in the well-known treatise by Kolmogoroff [53]. We shall begin by recalling some of the main definitions. Note that the word *additive*, when used in connection with sets or set functions, will always refer to a *finite or enumerable* sequence of sets.

Let $\omega$ denote a variable point in an entirely arbitrary space $\Omega$, and consider an additive class $C$ of sets in $\Omega$, such that the whole space $\Omega$ itself is a member of $C$. Further, let $P(S)$ be an additive set function, defined for all sets $S$ belonging to the class $C$, and suppose that

$$P(S) \geqq 0 \text{ for all } S \text{ in } C,$$

$$P(\Omega) = 1.$$

We shall then say that $P(S)$ is a *probability measure*, which defines a *probability distribution in* $\Omega$. For any set $S$ in $C$, the quantity $P(S)$ is called the *probability* of the *event* expressed by the relation $\omega \subset S$, i.e. the event that the variable point $\omega$ takes a value belonging to $S$. Accordingly we write

$$P(S) = P(\omega \subset S).$$

Suppose now that $\omega' = g(\omega)$ is a function of the variable point $\omega$, defined throughout the space $\Omega$, the values $\omega'$ being points of another arbitrary space $\Omega'$. Let $S'$ be a set in $\Omega'$ and denote by $S$ the set of all points $\omega$ such that $\omega' = g(\omega)$ belongs to $S'$. Whenever $S$ belongs to $C$, we define a set function $P'(S')$ by writing

$$P'(S') = P(S).$$

It is then easy to see that $P'(S')$ is defined for all $S'$ belonging to a certain additive class $C'$ in the new space $\Omega'$, and that $P'(S')$ is a probability measure in $\Omega'$, such that $P'(S')$ signifies the probability of the event $\omega' \subset S'$ (which is equivalent to $\omega \subset S$). We shall say that $P'(S')$ is attached to the probability distribution in $\Omega'$ which is *induced* by the given distribution in $\Omega$ and the function $\omega' = g(\omega)$.

**3. Random variables.** Consider in particular the case when $\omega'$ is a real number $\xi$, such that $\xi = g(\omega)$ is a real-valued $C$-measurable function of the argument $\omega$. Then $C'$ includes the class $B_1$ of all Borel sets $S'$ of the space $\Omega' = R_1$ of all real numbers, and we shall call $\xi$ a *one-dimensional real random variable.* The probability of the event $\xi \subset S'$ is uniquely defined for any Borel set $S'$ of $R_1$, as soon as the function

$$F(x) = P(\xi \leqq x)$$

is known for all real $x$. $F(x)$ is called the *distribution function* (*d.f.*) of the random variable $\xi$. If the function $\xi = g(\omega)$ is integrable over $\Omega$ with respect to the measure $P(S)$, we write

$$E\xi = \int_\Omega g(\omega)\, dP = \int_{-\infty}^{\infty} x\, dF(x),$$

and denote this expression as the *expectation* or *mean value* of the random variable $\xi$. Any real-valued $B$-measurable function $\eta = h(\xi)$ is also a random variable with the probability distribution induced by the original $\omega$-distribution and the function $\eta = h(g(\omega))$. If $\eta$ is integrable over $\Omega$ with respect to $P$, its mean value may be written in the form

$$E\eta = Eh(\xi) = \int_\Omega h(g(\omega))\, dP = \int_{-\infty}^{\infty} h(x)\, dF(x).$$

More generally, if $\omega' = (\xi_1, \cdots, \xi_n)$ is a point in an $n$-dimensional Euclidean space $R_n$, while $C'$ includes the class $B_n$ of all Borel sets of $R_n$, we are concerned with an *n-dimensional real random variable.* The distribution of this variable, which is also called the joint distribution of the $n$ one-dimensional variables $\xi_1, \cdots, \xi_n$, is uniquely defined, as soon as the joint d.f.

$$F(x_1, \cdots, x_n) = P(\xi_1 \leqq x_1, \cdots, \xi_n \leqq x_n)$$

is known for all real $x_1, \cdots, x_n$.

The variables $\xi_1, \cdots, \xi_n$ are said to be *independent,* if $F(x_1, \cdots, x_n) = F_1(x_1) \cdots F_n(x_n)$, where $F_\nu(x_\nu)$ is the d.f. of the variable $\xi_\nu$.

The extension to *complex random variables* is obvious. Suppose e.g. that $\xi = g(\omega)$ and $\eta = h(\omega)$ are two one-dimensional real variables, and consider the complex variable $\xi + i\eta = g(\omega) + ih(\omega)$. By definition, we identify the distribution of this variable with that of the two-dimensional real variable $(\xi, \eta)$, and we put

$$E(\xi + i\eta) = E\xi + iE\eta.$$

Joint distributions of several complex variables are introduced in a corresponding way.

**4. Characteristic functions.** If $\xi$ is a one-dimensional real random variable, the mean value

$$\varphi(z) = Ee^{iz\xi} = \int_{-\infty}^{\infty} e^{izx}\, dF(x)$$

exists for all real $z$, and we have

$$|\varphi(z)| \leqq 1, \qquad \varphi(0) = 1.$$

$\varphi(z)$ is called the *characteristic function* (*c.f.*) of the distribution corresponding to the variable $\xi$. The reciprocal formula (Lévy)

$$F(x) - F(y) = -\frac{1}{2\pi i} \lim_{Z\to\infty} \int_{-Z}^{Z} \frac{e^{-izx} - e^{-izy}}{z} \varphi(z)\, dz,$$

which holds for any continuity points $x$ and $y$ of $F$, shows that there is a one-one correspondence between the d.f. $F(x)$ and the c.f. $\varphi(z)$. As we shall see below, the c.f. provides a powerful analytical tool for operations with probability distributions.

When a complex-valued function $\varphi(z)$ of the real variable $z$ is given, it is often important to be able to decide whether $\varphi(z)$ is or is not the c.f. of some distribution. If we assume a priori that $\varphi(0) = 1$, each of the following conditions is necessary and sufficient for $\varphi(z)$ to be a c.f.

*A*. $\varphi(z)$ should be bounded and continuous for all $z$, and such that the integral

$$\int_0^A \int_0^A \varphi(z - u) e^{ix(z-u)}\, dz\, du$$

is real and non-negative for all real $x$ and all $A > 0$ (Cramér [11], in simplification of an earlier result due to Bochner, [4]).

*B*. There should exist a sequence of functions $\psi_1(z), \psi_2(z), \cdots$ such that

$$\varphi(z) = \lim_{n\to\infty} \int_{-\infty}^{\infty} \psi_n(x + z)\overline{\psi_n(x)}\, dx$$

holds uniformly in every finite $z$-interval (Khintchine, [45]).

These general theorems are not always easy to apply in practice. Among less general results which are more easily applicable, we mention the almost trivial fact that a function $\varphi(z)$ which near $z = 0$ is of the form $\varphi(z) = 1 + o(z^2)$ cannot be a c.f. unless $\varphi(z) = 1$ for all $z$, and the two following theorems:

1) An integral function $\varphi(z)$ of order $\gamma < 1$ can never be a c.f. (Lévy, [64]), and 2) an integral function $\varphi(z)$ of finite order $\gamma > 2$ cannot be a c.f. unless the convergence exponent of its zeros is equal to $\gamma$ (Marcinkiewicz, [72]). The latter result shows e.g. that no function of the form $e^{g(z)}$, where $g(z)$ is a polynomial of degree $> 2$, can be a c.f.

It would be highly desirable to obtain further results in this direction.

The c.f. of the joint distribution of $n$ real random variables $\xi_1, \cdots, \xi_n$ is the function $\varphi(z_1, \cdots, z_n)$ defined by the relation

$$\varphi(z_1, \cdots, z_n) = Ee^{i(z_1\xi_1+\cdots+z_n\xi_n)}.$$

Most of the above results for c.f. in one variable can be directly generalized to the multi-variable case.

**5. Random sequences and random functions.** Let $t$ be a variable point in an arbitrary space $\mathbf{T}$, and consider the space $\Omega$, where each point $\omega$ is a real-valued function $\omega = x(t)$ of the variable argument $t$. Let $t_1, \cdots, t_n$ be any finite set of distinct points $t$. The set of all functions $\omega = x(t)$ satisfying the inequalities

$$a_j < x(t_j) \leqq b_j, (j = 1, \cdots, n),$$

will be called an *interval* in the space $\Omega$. The Borel sets in $\Omega$ will be defined as the smallest additive class $B$ of sets in $\Omega$ containing all intervals.

Suppose now that, for any choice of $n$ and the $t_j$, the variables $x(t_1), \cdots, x(t_n)$ are random variables having a known $n$-dimensional joint distribution. If the family of all distributions corresponding in this way to finite sequences $t_1, \cdots, t_n$ satisfies certain obvious consistency conditions, a fundamental theorem due to Kolmogoroff asserts that this family determines a unique probability distribution in the space $\Omega$ of all functions $x(t)$. The corresponding probability

$$P(S) = P(x(t) \subset S)$$

is uniquely defined for all Borel sets $S$ of $\Omega$.

Consider in particular the case where $\mathbf{T}$ is the set of non-negative integers $t = 0, 1, 2, \cdots$. The space $\Omega$ then is the space of all sequences $(x_0, x_1, \cdots)$ of real numbers. As soon as the joint distribution of any finite number of variables $x_{\nu_1}, \cdots, x_{\nu_n}$ is defined, and these distributions are mutually consistent, it then follows that there is a unique probability distribution of the *random sequence* $(x_0, x_1, \cdots)$, the corresponding probability being defined for every Borel set of the space $\Omega$ of sequences. Similarly we may consider the doubly infinite sequence $(\cdots, x_{-1}, x_0, x_1, \cdots)$.

Consider further the more general case when $\mathbf{T}$ is any set of real numbers. Then $\Omega$ is the space of all real-valued functions $\omega = x(t)$ defined on the set $\mathbf{T}$, and as before the knowledge of the distributions for all finite sets of variables $x(t_1), \cdots, x(t_n)$ permits us to determine a probability distribution in the space $\Omega$ of *random functions* $x(t)$, the probability $P(S) = P(x(t) \subset S)$ being always defined for all Borel sets $S$ in $\Omega$.

The generalization of the above considerations to *complex-valued* random sequences and functions is immediate.

**6. Various modes of convergence.** Consider a sequence $F_1(x), F_2(x), \ldots$ of d.f:s, and let the corresponding c.f:s be $\varphi_1(t), \varphi_2(t), \cdots$. In order that $F_n(x)$

converge to a d.f. $F(x)$, in every continuity point of the latter, it is necessary and sufficient[1] that $\varphi_n(t)$ converge for every real $t$ to a limit $\varphi(t)$ which is continuous at $t = 0$. Then $\varphi(t)$ is the c.f. corresponding to the d.f. $F(x)$.

Further, let $x$ and $x_1, x_2, \cdots$ be complex-valued random variables, such that the random sequence $(x, x_1, x_2, \cdots)$ has a well defined distribution. We shall be concerned with various modes of convergence of $x_n$ to $x$.

*A*. When $P(|x_n - x| > \epsilon) \to 0$ as $n \to \infty$, for any $\epsilon > 0$, we shall say that *$x_n$ converges to $x$ in probability.*

*B*. When $E|x_n - x|^\gamma \to 0$, as $n \to \infty$, where $\gamma > 0$ is fixed, we shall say that $x_n$ converges to $x$ in the mean of order $\gamma$. Unless otherwise stated we shall in the sequel always consider the case $\gamma = 2$, and in this case we shall use the notation

$$\underset{n\to\infty}{\text{l.i.m.}}\, x_n = x.$$

*C*. When $P(\lim_{n\to\infty} x_n = x) = 1$, we shall say that $x_n$ converges with probability one, or *converges almost certainly to $x$.*

With respect to the last definition, we may remark that the set defined by the relation $\lim x_n = x$ is always a Borel set in the space of our random sequence, so that the probability of this relation is well defined. In fact, this probability is given by the expression

$$\lim_{m\to\infty} \lim_{n\to\infty} \lim_{p\to\infty} P\left(|x_\nu - x| < \frac{1}{m} \quad \text{for} \quad \nu = n, n+1, \cdots, n+p\right)$$

where the limit process applies to a probability attached to a Borel set in a finite number of dimensions. The case of almost certain convergence is precisely the case when this expression takes the value 1.

Convergence in the mean of any positive order, as well as almost certain convergence, both imply convergence in probability, which may be written symbolically $B \to A$ and $C \to A$. Between $B$ and $C$, there is no simple relation of this kind. Further, $A$ and $B$ both imply almost certain convergence for any partial sequence $x_{n_1}, x_{n_2}, \cdots$ such that the subscripts $n_k$ increase sufficiently rapidly with $k$.

## II. PROBLEMS CONNECTED WITH THE ADDITION OF INDEPENDENT VARIABLES

**7.** During the early development of the theory of probability, the majority of problems considered were connected with gambling. The gain of a player in a certain game may be regarded as a random variable, and his total gain in a

[1] As I have already stated in a paper published in 1938, there is an error in the statement of this theorem given in my Cambridge Tract [9] *Random Variables and Probability Distributions*. For the truth of the theorem, it is essential that $\varphi_n(t)$ should be supposed to converge to $\varphi(t)$ *for every real $t$*. However, in the particular case when the limit $\varphi(t)$ is analytic and regular in the vicinity of $t = 0$, it can be proved that it is sufficient to assume convergence in some interval $|t| < a$.

sequence of repetitions of the game is the sum of a number of independent variables, each of which represents the gain in a single performance of the game. Accordingly a great amount of work was devoted to the study of the probability distributions of such sums. A little later, problems of a similar type appeared in connection with the theory of errors of observation, when the total error was considered as the sum of a certain number of partial errors due to mutually independent causes. At first only particular cases were considered, but gradually general types of problems began to arise, and in the classical work of Laplace several results are given concerning the general problem to study the distribution of a sum

$$z_n = x_1 + \cdots + x_n$$

of independent variables, when the distributions of the $x_j$ are given. This problem may be regarded as the very starting point of a large number of those investigations by which the modern Theory of Probability was created. The efforts to prove certain statements of Laplace, and to extend his results further in various directions, have largely contributed to the introduction of rigorous foundations of the subject, and to the development of the analytical methods. At the same time, more general types of problems have developed from the original problem, and the number and importance of practical applications have been steadily increasing.

**8. Composition of distributions.** Let $x_1$ and $x_2$ be two independent variables, with the d.f.'s $F_1$ and $F_2$, and the c.f.'s $\varphi_1$ and $\varphi_2$, and let the sum $x_1 + x_2$ have the d.f. $F$ and the c.f. $\varphi$. Then

$$F(x) = \int_{-\infty}^{\infty} F_1(x - y)\, dF_2(y) = \int_{-\infty}^{\infty} F_2(x - y)\, dF_1(y).$$

We shall say that $F$ is the *composition* of $F_1$ and $F_2$, and write this as a symbolical multiplication:

$$F = F_1 * F_2 = F_2 * F_1 .$$

To this symbolical multiplication of the d.f:s corresponds a real multiplication of the c.f.'s:

$$\varphi(z) = \varphi_1(z)\varphi_2(z).$$

The operation of composition is both commutative and associative, so that any symbolical product $F = F_1 * F_2 \cdots * F_n$ is uniquely defined and independent of the order of the components. When at least one of the components is continuous (absolutely continuous), the same holds for the composite, and in many cases it is true that the composite is at least as regular as the most regular of the components (Lévy, [58], [63], etc.). However, this general statement does not hold generally, as is shown by an interesting example due to Raikov, [77], where $F_1$ and $F_2$ are integral analytic functions, while the composite $F = F_1 * F_2$ is not regular at the origin.

It seems to be an important unsolved problem to find convenient restrictions ensuring the validity of the above statements of the "smoothing effect" of the operation of composition.

When $F = F_1 * F_2$, we may say that $F$ is "divisible" by each component $F_1$ and $F_2$, and it seems natural to try to develop a theory of symbolical factorization for d.f.'s. In this connection, it is important to note that symbolical division is not unique. In fact, Khintchine has shown by an example that it is possible to find the d.f.'s $F$, $F_1$, $F_2$, and $F_3$ such that [33]

$$F = F_1 * F_2 = F_1 * F_3,$$

p·17

while $F_2 \neq F_3$. Another fundamental problem belonging to this order of ideas is to decide whether a given d.f. $F$ is decomposable or not. $F$ is called decomposable, if there is at least one representation of the form $F = F_1 * F_2$, where each component $F_\nu$ has more than one point of increase. So far, this problem has only been solved in very special cases, and the general problem still remains open for research. A particular case of some interest would be to know if there exists an absolutely continuous and indecomposable d.f., such that $F(a) = 0$ and $F(b) = 1$ for some finite $a$ and $b$.

As soon as we restrict ourselves to certain special classes of distributions, it is possible to reach results of a more definite character concerning the factorization problems. Some results of this type will be considered below.

**9. Closed families of distributions.** The fact that certain families of distributions are closed with respect to the operation of composition has played an important part in many applications. If $F_1$ and $F_2$ belong to a family of this character, so does the symbolical product $F = F_1 * F_2$. We first give some simple examples of such families.

*The normal distribution.* The d.f. $F$ has the form $F = \phi\left(\frac{x - m}{\sigma}\right)$, where $\sigma > 0$, and

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(t^2/2)}\, dt.$$

The c.f. corresponding to $F$ is $e^{miz - \frac{1}{2}\sigma^2 z^2}$, and it follows that for any real $m_1$, $m_2$ and any positive $\sigma_1$, $\sigma_2$ we have

$$\phi\left(\frac{x - m_1}{\sigma_1}\right) * \phi\left(\frac{x - m_2}{\sigma_2}\right) = \phi\left(\frac{x - m}{\sigma}\right),$$

where

$$m = m_1 + m_2, \qquad \sigma^2 = \sigma_1^2 + \sigma_2^2.$$

*The Poisson distribution.* Here the d.f. is $F = F(x; \lambda, m, a)$ where $\lambda > 0$, $a \neq 0$, and $F$ is a step-function with a jump equal to $\frac{\lambda^\nu}{\nu!} e^{-\lambda}$ in the point $x = m + \nu a$, where $\nu = 0, 1, \cdots$. The corresponding c.f. is $e^{miz + \lambda(e^{aiz} - 1)}$, and it follows that for any fixed $a$ we have

$$F(x; \lambda_1, m_1, a) * F(x; \lambda_2, m_2, a) = F(x; \lambda_1 + \lambda_2, m_1 + m_2, a).$$

*The Pearson Type III distribution.* $F = F(x; \alpha, \lambda) = \frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^x t^{\lambda-1} e^{-\alpha t}\, dt$, $(x > 0)$. The corresponding c.f. is $\left(1 - \frac{iz}{\alpha}\right)^{-\lambda}$, and for any fixed $\alpha > 0$ and any positive $\lambda_1$ and $\lambda_2$ we have

$$F(x; \alpha, \lambda_1) * F(x; \alpha, \lambda_2) = F(x; \alpha, \lambda_1 + \lambda_2).$$

*Stable distributions.* We shall say that a closed family is stable, when all its members are of the form $F(ax + b)$, where $F$ is a d.f., while $a > 0$ and $b$ are constants. Obviously the normal family is an example of a stable family. It has been shown by Lévy and Khintchine [49], that a d.f. $F(x)$ generates a stable family when and only when the logarithm of its c.f. is of the form

$$\log \varphi(z) = \beta iz - \gamma \,|\, z \,|^\alpha \left(1 + i\delta \frac{z}{|\, z \,|}\, \omega\right), \tag{9.1}$$

where $\alpha, \beta, \gamma, \delta$ are real constants such that

$$0 < \alpha \leqq 2, \qquad \gamma > 0, \qquad |\, \delta \,| \leqq 1,$$

while

$$\omega = \begin{array}{ll} tg\, \frac{\alpha\pi}{2} & \text{for} \quad \alpha \neq 1 \\ \frac{2}{\pi} \log |\, z \,| & \text{for} \quad \alpha = 1. \end{array}$$

For $\alpha = 2$ we obtain the normal family.

A more general and very important closed family is the family $I$ of *infinitely divisible distributions.* A d.f. $F$ belongs to $I$ if to every $n = 1, 2, \cdots$ there exists a d.f. $G$ such that $F = G^{[n]}$, where $G^{[n]}$ denotes the symbolical $n$th power of $G$. Obviously the family $I$ is a closed family which contains all the families mentioned above. Lévy [60], [63], has shown that $F$ is infinitely divisible when and only when the logarithm of its c.f. is of the form

$$\begin{aligned} \log \varphi(z) = \beta iz - \gamma z^2 + \int_{-\infty}^0 \left(e^{izu} - 1 - \frac{izu}{1 + u^2}\right) dM(u) \\ + \int_0^\infty \left(e^{izu} - 1 - \frac{izu}{1 + u^2}\right) dN(u), \end{aligned} \tag{9.2}$$

where $\beta$ and $\gamma > 0$ are real constants, while $M(u)$ and $N(u)$ are non-decreasing functions such that

$$M(-\infty) = N(+\infty) = 0,$$

$$\int_{-a}^0 u^2\, dM(u) < \infty \quad \text{and} \quad \int_0^a u^2\, dN(u) < \infty$$

for any finite $a > 0$. When $M$ and $N$ reduce to zero, we obtain the normal family. When $\gamma = 0$ and one of the functions $M$ and $N$ reduces to zero, while the other is a step-function with a single jump equal to $\lambda$ at the point $x = a$, we obtain a Poisson family. Generally, it follows from (9.2) that any infinitely divisible distribution may be regarded as a product of a normal distribution and a finite, enumerable or continuous set of Poisson distributions.

The representation of $\log \varphi(z)$ in the form (9.2) is unique. It follows that the problem of finding all possible factorizations of an infinitely divisible d.f. $F$ can be completely solved, as long as we restrict ourselves to factors which are themselves infinitely divisible. In fact, in order that

$$F = F_1 * F_2,$$

where all three d.f.'s belong to $I$, it is necessary and sufficient that the logarithms of the corresponding c.f.'s should be of the form (9.2), with

$$\beta = \beta_1 + \beta_2, \qquad \gamma = \gamma_1 + \gamma_2,$$
$$M = M_1 + M_2, \qquad N = N_1 + N_2.$$

In the two simple cases of the normal and the Poisson distributions, the decompositions obtained in this way remain the only possible, even if we remove the restriction that the factors should belong to $I$. Thus in any factorization of a normal distribution, all factors are normal (Cramér, [8]), while in any factorization of a Poisson distribution, all factors belong to the Poisson family (Raikov, [75]). For the type III distribution, and the non-normal stable distributions, however, the corresponding property does not hold.

In some cases, an infinitely divisible distribution may be represented as a product of indecomposable distributions, or as a product of an indecomposable distribution and another infinitely divisible distribution. The results so far obtained in this direction (Lévy, [63], [64], Khintchine, [46], [47]; Raikov, [76]) are all concerned with more or less particular cases, and the general factorization problem for infinitely divisible distributions still remains unsolved. A particular case of some interest would be the case when the functions $M$ and $N$ are both absolutely continuous. There does not seem to have been given any example of this type, where a factor not belonging to $I$ may occur.[2]

Finally we mention a general theorem due to Khintchine, [46], which asserts that an arbitrary d.f. $F$ may be represented in one of the forms

$$F = G, \qquad F = H \text{ or } F = G * H,$$

where $G$ is infinitely divisible, while $H$ is a finite or infinite product of indecomposable factors. This seems to be practically the only result so far known concerning the factorization of a general distribution.

A certain number of the results mentioned above have been generalized to multi-dimensional distributions.

[2] While the present paper was being printed, I have proved that such factors do occur, as soon as at least one of the derivatives $M'$ and $N'$ is bounded away from zero in some interval $(-a, 0)$ or $(0, a)$.

**10. The Laws of large numbers.** In modern terminology, the classical Bernoulli theorem may be expressed in the following way. Let $x_1, x_2, \cdots$ be a sequence of independent variables, such that each $x_\nu$ may only assume the values 1 and 0, the corresponding probabilities being $p$ and $q = 1 - p$. Then the arithmetic mean

$$\frac{z_n}{n} = \frac{x_1 + \cdots + x_n}{n} \tag{10.1}$$

converges in probability to $p$, as $n \to \infty$.

Both classical and modern authors have laid down much work on the generalization of this simple result in various directions. Generally, we shall say that a sequence of random variables $x_1, x_2, \cdots$ satisfies the *Weak Law of Large Numbers* if there exist two sequences of constants $a_1, a_2, \cdots$ and $b_1, b_2, \cdots$, such that $a_n > 0$, and

$$\frac{z_n - b_n}{a_n} = \frac{x_1 + \cdots + x_n - b_n}{a_n}$$

converges in probability to zero.

Let $x_1, x_2 \cdots$ be independent variables, such that $x_\nu$ has the d.f. $F_\nu(x)$. It has been shown by Feller [27] that *for any given sequence* $a_1, a_2, \cdots$, *the conditions*

$$\begin{aligned} \sum_{\nu=1}^{n} \int_{|x|>a_n} dF_\nu(x) &= o(1), \\ \sum_{\nu=1}^{n} \int_{|x|<a_n} x^2\, dF_\nu(x) &= o(a_n^2), \end{aligned} \tag{10.2}$$

*are sufficient for the validity of the weak law of large numbers, and that the corresponding sequence* $b_1, b_2, \cdots$ *can be defined by*

$$b_n = \sum_{\nu=1}^{n} \int_{|x|<a_n} x\, dF_\nu(x).$$

*When there is a constant* $c > 0$ *such that for all* $\nu$

$$F_\nu(+0) > c, \qquad F_\nu(-0) < 1 - c, \tag{10.3}$$

*the conditions are also necessary.* This theorem contains as particular cases all previously known results in this direction. A simple $NS$ condition for the existence of at least one sequence $a_1, a_2, \cdots$ such that 10.2 holds does not seem to be known.

When the weak law is satisfied, this means that, for any given $\epsilon > 0$ and *for any fixed large* $n$, there is a probability very near to 1 that the sum $z_n = x_1 + \cdots + x_n$ will fall between the limits $b_n \pm \epsilon a_n$. The more stringent condition that, with a probability tending to 1 as $n \to \propto$, $z_\nu$ will fall between the limits $b_\nu \pm \epsilon a_\nu$ *for all values of* $\nu \geqq n$ is equivalent to the condition that $\frac{z_n - b_n}{a_n}$ converges *almost certainly* to zero. When this holds, we shall say that the variables $x_\nu$ satisfy the *Strong Law of Large Numbers.* The most important result so far known in this connection is concerned with the case $a_n = n$, and is expressed by the following theorem (Kolmogoroff, [52], [55]):

*When the $x_\nu$ are independent and (10.3) holds, a sufficient condition for the validity of the strong law with $a_n = n$ consists in the simultaneous convergence of the two series*

$$\Sigma \int_{|x| > n} dF_n(x) \qquad \textit{and} \qquad \Sigma \frac{1}{n^2} \int_{|x| < n} x^2 \, dF_n(x).$$

Some improved conditions of this type have been given by Marcinkiewicz and Zygmund, [73], but the problem of finding a *NS* condition for the strong law is still unsolved, even in the case $a_n = n$.

Important generalizations of the laws of large numbers to cases when the $x_\nu$ are not assumed to be independent have been given i.a. by Khintchine [44], Lévy [62], [63] and Loève [67].

**11. The central limit theorem and allied theorems.** It was already known to De Moivre that, in the case 10.1 of the Bernoulli distribution, the d.f. of the normalized sum

$$\frac{x_1 + \cdots + x_n - np}{\sqrt{npq}}$$

tends, as $n \to \infty$, to the normal d.f. $\phi(x)$. Considerably more general results in this direction were stated by Laplace. After a long series of more or less successful attempts, a rigorous proof of the main statements of Laplace was given in 1901 by Liapouncff, [65]. More general cases were later considered i.a. by Lindeberg [66], Lévy [61], [63], Khintchine [43] and Feller, [25]. The following final form of the *Central Limit Theorem* is due to Feller.

Consider the expression

$$(11.1) \qquad u_n = \frac{z_n - b_n}{a_n} = \frac{x_1 + \cdots + x_n - b_n}{a_n},$$

where the $x_\nu$ are independent variables. We shall say that the $x_\nu$ obey the central limit law, if the sequences $\{a_\nu\}$ and $\{b_\nu\}$ can be found such that the d.f. of $u_n$ tends to $\phi(x)$ as $n \to \infty$. In order to avoid unnecessary complications, we shall restrict ourselves to sequences $\{a_\nu\}$ such that

$$a_\nu \to +\infty, \qquad \frac{a_{\nu+1}}{a_\nu} \to 1,$$

and we shall assume that the conditions (10.3) are satisfied. Then Feller's theorem runs as follows:

*The independent variables $x_1, x_2, \cdots$ obey the central limit law if, and only if, there exists a sequence $q_n \rightarrow \infty$ such that simultaneously*

$$(11.2)\qquad \begin{aligned} &\sum_{\nu=1}^{n} \int_{|x|>q_n} dF_\nu(x) \rightarrow 0, \\ &\frac{1}{q_n^2} \sum_{\nu=1}^{n} \int_{|x|<q_n} x^2\, dF_\nu(x) \rightarrow \infty. \end{aligned}$$

*When these conditions are satisfied, explicit expressions for the $a_n$ and $b_n$ can be obtained.*

Feller's theorem gives a complete solution of the problem. However, we might still try to express in a more direct way the condition that the $q_n$ should exist. We may also ask what happens when the conditions (11.2) are not satisfied. Some particular cases of the latter question will be considered below. However, very few general results are known in this direction.

The central limit theorem has been extended in various directions. Bernstein [3], Lévy [62], [63], Loève [67] and others have considered cases where the $x_\nu$ are not assumed to be independent. Important results have been reached but still much remains to be done.

On the other hand, several authors have considered symmetrical functions, other than sums, of $n$ independent random variables. The problem of investigating the asymptotic behaviour of the distributions of such functions, as $n$ tends to infinity, is of great importance in the theory of statistical sampling distributions. It is known (c.f. e.g. Cramér, [15]) that under certain general regularity conditions there exists a normal limiting distribution. However, it is also known that it is possible to give examples of particular functions (such as e.g. the function which is equal to the largest of the $n$ variables), where there exist limiting distributions which are non-normal. The conditions under which this phenomenon may occur seem to deserve further study.

A further problem belonging to the same order of ideas is to find a closer asymptotic representation of the d.f. of the standardized sum $z_n$ than that provided by the normal function $\phi(x)$. Consider e.g. the simple case when the $x_\nu$ are independent variables all having the same d.f. $F(x)$ with a finite mean $m$, a finite variance $\sigma^2$, and finite moments up to a certain order $k \geqq 3$. Let $G_n(x)$ be the d.f. of the variable

$$\frac{x_1 + \cdots + x_n - nm}{\sigma\sqrt{n}}.$$

It then follows from a theorem of Cramér [5], [9] that, as soon as the d.f. $F(x)$ contains an absolutely continuous component, there is an asymptotic expansion

$$(11.3)\qquad G_n(x) = \phi(x) + \sum_{\nu=1}^{k-3} \frac{p_\nu(x)}{n^{\nu/2}} e^{-x^2/2} + O(n^{-(k-2)/2}),$$

where the constant implied by the $O$ is independent of $n$ and $x$. Cramér has also given similar expansions in more general cases, and his results have been further extended by P. L. Hsu [39], who deduces analogous expansions also for other functions than sums. The most general conditions under which expansions of this type exist are still unknown.

It follows from (11.3) that the difference $G_n(x) - \phi(x)$ is, for any fixed $x$, of the order $n^{-\frac{1}{2}}$ as $n \rightarrow \infty$. It is often important to know the asymptotic behaviour of $G_n(x)$ when $n$ and $x$ increase simultaneously, and in that case (11.3) yields only a trivial result. This case has been investigated by Cramér [10], and Feller [29], and the results so far obtained permit important applications to the so called law of the iterated logarithm (cf. below). However, it seems likely that similar results may be obtained in considerably more general cases than those hitherto investigated.

A further interesting type of problems belonging to this order of ideas may be approached in the following way. Consider the variables (11.1) in the particular case when $x_1, x_2, \cdots$ are independent variables all having the same d.f. $F(x)$. When the $a_n$ and $b_n$ can be found such that the d.f. of the normalized sum $u_n$ tends to $\phi(x)$, we shall say that $F$ belongs to the *domain of attraction* of the normal law. Feller's theorem gives a $NS$ condition that this should be so. Now when this condition is not satisfied, it may still occur that the $a_n$ and $b_n$ can be so chosen that the d.f. of $u_n$ tends to a limiting d.f. $\Psi(x)$, which is necessarily different from $\phi(x)$. Then it is easily seen that $\Psi(x)$ must be a stable distribution, with its c.f. defined by (9.1), and it is natural to say that $F$ belongs to the domain of attraction of $\Psi$. $NS$ and sufficient conditions that this should hold have been given by Doeblin [16], and Gnedenko [34]. When the $a_n$ and $b_n$ cannot be found such that the d.f. of the normal sum $u_n$ converges to a limit, it may still be possible to obtain a limiting d.f. by considering only a partial sequence $u_{n_1}, u_{n_2}, \cdots$. Khintchine [47] has proved the interesting theorem that the totality of limiting d.f.'s that may be obtained in this way coincides with the class of infinitely divisible d.f.'s defined by (9.2). There are also further results in the same direction given by Bawly [2], Khintchine [44], Lévy, [61]-[63], and Gnedenko, [35].

**12. The law of the iterated logarithm.** Consider a sequence of independent variables $x_1, x_2, \cdots$, such that the mean $Ex_n = 0$ for all $n$, while the variances $Ex_n^2 = \sigma_n^2$ are finite. Put $s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2$, and suppose that the variables obey the central limit law with $a_n = s_n$, $b_n = 0$. (In particular this will be the case when all $x_n$ have the same distribution.) For any function $\psi(n)$ tending to infinity with $n$ we then have

$$(12.1) \qquad \lim_{n \to \infty} P(| z_n | > s_n \psi(n)) = 0.$$

On the other hand, if $\psi(n)$ tends to a finite limit $> 0$, the same probability has a positive limit.

It seems natural to consider the relation within the brackets in (12.1) not only for a single large value of $n$, but to require the probability that this relation holds simultaneously for *an infinite number of values of n*. The development of this problem has led to the so called law of the iterated logarithm.

We shall in this respect use the following terminology due to Lévy. A non-decreasing positive function $\psi(n)$ will be said to belong to the *lower class* with respect to the variables $x_n$ if, with a probability equal to one, there are infinitely many $n$ such that

$$| z_n | > s_n \psi(n).$$

On the other hand, $\psi(n)$ will be said to belong to the *upper class* if the probability of the same property is equal to zero.

Every $\psi(n)$ belongs to one of these two classes. This is a special case of the so called *null-or-one law*: if $S$ is a Borel set in the space of the independent random variables $x_1, x_2, \cdots$, such that any two points differing at most in a finite number of coordinates either both belong to $S$ or both belong to the complementary set, then $P(S)$ can only assume the values 0 or 1.

It was proved by Kolmogoroff [51] that, subject to certain restrictions, the function

$$\psi(n) = \sqrt{c \log \log s_n}$$

belongs to the lower class for any $c < 2$, and to the upper class for any $c > 2$, which may be expressed by the relation

$$P\left(\limsup \frac{z_n}{s_n \sqrt{2 \log \log s_n}} = 1\right) = 1. \tag{12.1}$$

More general results were proved by Feller [30], who proved i.a. that, subject to certain restrictions, $\psi(n)$ belongs to the lower or upper class according as

$$\Sigma \frac{\sigma_n^2}{s_n^2} \psi(n) e^{-(\psi^2(n)/2)} \tag{12.2}$$

is divergent or convergent (in certain special cases, this had been previously found by Kolmogoroff and Erdös [24]. Feller also proved a more complicated result, which contains the above as a particular case, and from which it follows that the simple criterion (12.2) no longer holds when the restrictions imposed in its proof are removed.

**13. Convergence of series.** For any sequence of random variables $x_n$, the probability

$$P\left(\sum_1^\infty x_n \text{ converges}\right)$$

has a uniquely determined value. When the $x_n$ are independent, it follows from the null-or-one law that this probability is either 0 or 1. By a theorem of Khintchine and Kolmogoroff [48], the value 1 is assumed when and only when the three series

$$\sum_n \int_{|x_n|>1} dF_n, \qquad \sum_n Ey_n, \qquad \sum_n \sigma^2 y_n$$

are convergent, where

$$y_n = \begin{matrix} x_n & \text{when} & |x_n| \leqq 1. \\ 0 & \text{when} & |x_n| > 1. \end{matrix}$$

For the case when the $x_n$ are not assumed to be independent, various results have been given by Lévy [63] and others, but our knowledge of the properties of these series is still not very advanced.

**14. Generalizations.** In several instances it has been pointed out above that the results concerning sums of independent variables may, to a certain extent, be extended to cases when the variables are not independent. Generally the independence condition has then to be replaced by some condition restricting the degree of dependence. Results of this type were first give by Bernstein [3], and then in more general cases by Lévy [62], [63], and Loève [67]. However, this field has so far only been very incompletely explored.

Similar remarks apply to the generalization of the various theorems quoted above to cases of variables and distributions in more than one dimension.

## III. STOCHASTIC PROCESSES

**15.** The theory of random variables in a finite number of dimensions is able to deal adequately with practically all problems considered in classical probability theory. However, during the early years of the present century, there appeared in the applications various problems, where it proved necessary to consider probability relations bearing on infinite sequences of numbers, or even on functions of a continuous variable.

The mathematical set-up required for the study of such problems involves the introduction of probability distributions in spaces of random sequences or random functions (cf. 5 above). Generally, any process in nature which can be analyzed in terms of probability distributions in spaces of these types will be called a *stochastic process*. It is convenient to apply this name also to the probability distribution used for the study of the process. We shall thus say, e.g., that a certain random function $x(t)$ is attached to the stochastic process which is defined by the probability distribution of $x(t)$. In the majority of applications, the variable $t$ will represent the time, and we shall often use a terminology directly referring to this case. However, there are also other types of problems in the applications ($t$ may e.g. be a spatial variable in an arbitrary number of dimensions), and it is obvious that the purely mathematical problems connected with these classes of probability distributions will have to be considered quite independently of any concrete interpretation of the variable $t$ or the funcion $x(t)$.

A well-known example of this type of problems is afforded by the Brownian movement. Let $x(t)$ be the abscissa at the time $t$ of a small particle immersed in a liquid, and subject to molecular impacts. In every instant, the quantity $x(t)$ receives a random impulse, and the problem arises to study the behaviour of $x(t)$. According as we are content to consider $x(t)$ for a discrete sequence of $t$-points, say for $t = 0, 1, 2, \cdots$, or we wish to consider all positive values of $t$, we shall then have to introduce a probability distribution in the space of the random sequence $x(0)$, $x(1)$, $\cdots$, or in the space of the random function $x(t)$, where $t > 0$. We may then discuss such questions as the distribution of $x(t)$ for a given value of $t$, the joint and conditional distributions of $x(t)$ for two or more values of $t$, and, in the case of a continuous variable $t$, continuity, differentiability and other similar properties of the random function $x(t)$.

Wiener [82], [83] (cf. also Paley and Wiener [74]) was the first to give a rigorous treatment of this process. He proved in 1923 that it is possible to define a probability distribution in a suitably restricted functional space, such that the increment $\Delta x(t) = x(t + \Delta t) - x(t)$ is independent of $x(t)$ for any $\Delta t > 0$. With a probability equal to 1, the function $x(t)$ is continuous for all $t > 0$, and for any fixed $t > 0$, the random variable $x(t)$ is normally distributed.

Another example of stochastic processes studied at this stage occurs in the theory of risk of an insurance company. Let $x(t)$ denote the total amount of claims up to the time $t$ in a certain insurance company. As in the case of the Brownian movement, it may seem natural to assume that the increment $\Delta x(t)$ is independent of $x(t)$. On the other hand, $x(t)$ is in this case an essentially discontinuous function, which is never decreasing, and increases only by jumps of varying magnitudes occurring for certain discrete values of $t$, which are not a priori known. Processes of this type were studied by F. Lundberg [69], [70], H. Cramér [6] and others.

Further examples of particular processes were discussed in connection with various applications, but no general theory of the subject existed until 1931, when Kolmogoroff published a basic paper [53] dealing with the class of stochastic processes which will here be denoted as Markoff processes (Kolmogoroff uses the term "stochastically definite processes"), of which the two examples mentioned above form particular cases. The theory of this class of processes was further developed by Feller [26], [28]. In 1934, Khintchine [42] introduced another important class of processes known as stationary processes. From 1937, the general theory of the subject was subjected to a penetrating analysis in a series of important works by Doob [18]–[22].[3]

**16. Probability distributions in functional spaces.** We have seen in 5 above how a probability distribution in the space of all functions $x(t)$ may be defined, when $t$ varies in an arbitrary space T. Generally, we shall here content ourselves to consider the cases when T is the set of all real numbers, or the set of all non-negative real numbers. Most results obtained for these cases will be readily generalized to cases when $t$ varies in a Euclidean space of a finite number of dimensions. On the other hand, when T is enumerable, say consisting of the points $t = 0, \pm 1, \pm 2, \cdots$, so that we are concerned with a random sequence $x(0)$, $x(\pm 1)$, $\cdots$, the results for the continuous case will generally hold and assume a simpler form which will not be particularly stated here.

[3] A further interesting paper by Doob has appeared while the present paper was being printed: "Probability in function space", *Bull. Amer. Math. Soc.*, Vol. 53 (1947), pp. 15–30.

The case when T is a space of an infinite number of dimensions does not seem to have been considered so far.

In the present paragraph, it will be convenient to assume the function $x(t)$ to be real-valued, but the generalization to a complex-valued $x(t)$ requires only obvious modifications. In the sequel we shall sometimes consider the real-valued and sometimes the complex-valued case, according as the occasion requires.

Let now $X$ be the space of all real-valued functions $x(t)$ of the real variable $t$, where $-\infty < t < \infty$. According to 5, a probability measure $P(S)$ is uniquely defined for all Borel sets $S$ in $X$ by means of the family of joint distributions of all finite sequences $x(t_1), \cdots, x(t_n)$. In fact, $P(S)$ can be defined for a more general class of sets than the Borel sets. For any set $S$ in $X$, we may define an outer $P$-measure $\overline{P}(S)$ as the lower bound of $P(Z)$ for all sums $Z$ of finite or enumerable sequences of intervals, such that $S \subset Z$. Further, the inner $P$-measure $\underline{P}(S)$is defined by the relation $\underline{P}(S) = 1 - \overline{P}(X - S)$. When the outer and inner measures are equal, $S$ is called $P$-measurable, and $P(S)$ is defined as their common value. Any $P$-measurable set differs from a Borelset by a set of $P$-measure zero.

In many cases, this definition will be sufficient for an adequate treatment of the problems that we wish to consider. However, in other cases we encounter certain characteristic difficulties, which make it desirable to consider the possibility of amending the basic definition. Thus it often occurs that we are interested in the probability that the random function $x(t)$ satisfies certain regularity conditions in a non-enumerable set of points $t$. We may, e.g., wish to consider the probability that $x(t)$ is continuous for all $t$, that $x(t)$ should be Lebesque-measurable for all $t$, that $x(t) \leqq k$ for all $t$, etc. Let $S$ denote the set of all functions satisfying a condition of this type. It can then be shown that the inner measure $\underline{P}(S)$ is always equal to zero so that $S$ is never measurable, except in the (usually trivial) case when $P(S) = 0$.

Consequently many interesting probabilities are left undetermined by the general definition of a probability distribution in $X$ given above. The possibility of modifying the definition so as to enable us to study probabilities of this type has been thoroughly investigated by Doob [18]. He considers a subspace $X_0$ of the general functional space $X$, where $X_0$ is chosen so as to contain only, or almost only, "desirable" functions, i.e. functions satisfying such regularity conditions as seem natural with respect to the problem under investigation. We start from a given probability measure $P(S)$ in $X$, and ask if it is possible to define a probability measure in the restricted space $X_0$, which corresponds in some natural way to the given distribution in $X$. Let $S_0$ be a set in $X_0$, and suppose that it is possible to find a $P$-measurable set $S$ in $X$ such that $SX_0 = S_0$. According to Doob, a probability measure $P_0$ in $X_0$ is then uniquely defined by the relation

$$P_0(S_0) = P(S)$$

if and only if the condition

$$\overline{P}(X_0) = 1$$

is satisfied.

The problem is thus reduced to finding a subspace $X_0$ of outer $P$-measure 1, such that $X_0$ contains only functions of sufficiently regular behaviour. When this can be done, we can restrict ourselves to consider only functions $x(t)$ belonging to $X_0$, the probability distribution in this space being defined by the measure $P_0$. We shall then say that $x(t)$ is a random function, attached to a stochastic process with the restricted space $X_0$. Doob has obtained a great number of interesting results in this connection, e.g. with respect to the problem of choosing $X_0$ such that it contains almost only Lebesque-measurable functions, or such that the probability of the relation $x(t) \leqq k$ has a well-defined value for all $k$. In particular he has shown that the last problem can be solved for any given $P$-measure. However, our knowledge of the various possibilities which exist with respect to the choice of $X_0$ is still very incomplete, and it seems likely that further important results may be reached along this line of research.

An alternative method of introducing probability distributions in functional spaces has been used by Wiener [82], [83], (cf. also Paley and Wiener, [74]). Consider a given probability measure $\Pi$ in an arbitrary space $\Omega$, defined for all sets $\Sigma$ of an additive class $C$. Let $x(t, \omega)$ denote a function (real- or complex-valued, as the case may be) of the arguments $t$ (real) and $\omega$ (point in $\Omega$), such that $x(t, \omega)$ for every fixed $t$ becomes a $C$-measurable function of $\omega$. On the other hand, when $\omega$ is fixed, $x(t, \omega) = x(t)$ reduces to a function of the real variable $t$. Let $X_0$ denote the set of all functions $x(t)$ corresponding in this way to points of $\Omega$. Further, let $S_0 = SX_0$, where $S$ is a Borel set in $X$, and let $\Sigma$ denote the set of all points $\omega$ such that $x(t, \omega) \subset S_0$. Then $\Sigma$ belongs to $C$, and a probability measure $P_0$ in the functional space $X_0$ is uniquely defined by the relation

$$P_0(S_0) = \Pi(\Sigma). \tag{16.1}$$

The relations between the two modes of definition have been discussed by Doob and Ambrose [23] who have shown that they are largely equivalent. However, it seems likely that in particular problems the one or the other procedure may sometimes be the more advantageous, and further investigations on this subject seem desirable.

**17. Processes with a finite mean square.** Consider a stochastic process defined by a probability measure $P(S)$ in the space $X$ of all complex-valued functions $x(t)$ of the real variable $t$. For any fixed $t_0$, the random variable $x(t_0)$ is then a complex-valued function of the variable point $x(t)$ in the space $X$, i.e. a point $Q_{t_0}$ in the space $\Omega$ of all complex-valued functions defined on $X$. When $t_0$ varies, the point $Q_{t_0}$ describes a "curve" in $\Omega$, which then corresponds to our stochastic process.

Suppose, in particular, that the mean square

$$E \mid x(t) \mid^2 = \int_X \mid x(t) \mid^2 dP$$

is finite for any fixed value of $t$. This implies that for fixed $t$ the function $x(t)$ belongs to $L_2$ over $X$, relative to the probability measure $P$. The random variable $x(t)$ may then be regarded as an element of the Hilbert space $H$ of all complex-valued functions $f$ belonging to $L_2$ over $X$, the inner product $(f, g)$ of two elements $f$ and $g$ being defined by the relation

$$(f, g) = \int_X f\bar{g} \, dP = E(f\bar{g}).$$

The stochastic process to which $x(t)$ is attached then corresponds to a "curve" in $H$ (Kolmogoroff, [56], [57]), so that the well-known theory of Hilbert space is available for the study of the process. In particular, convergence in the usual metric of Hilbert space is equivalent to convergence in the mean of order 2 for random variables.

Let $H_x$ be the smallest closed linear subspace of $H$ which contains all elements of the form $a_1 x(t_1) + \cdots + a_n x(t_n)$. If the *covariance function*

$$r(t, u) = (x(t), x(u)) = E(x(t)\overline{x(u)})$$

is continuous for all real values of $t$ and $u$, then $x(t) \rightarrow x(t_0)$ in the mean, as $t \rightarrow t_0$, and we shall say that the process $x(t)$ is *continuous*. For any continuous process, $H_x$ is separable. When $g(t)$ is a continuous non-random function of $t$, and $x(t)$ is attached to a continuous stochastic process, the Riemann-Darboux sums formally associated with the integral

$$\int_a^b g(t)x(t) \, dt$$

are easily shown to tend to a limit $y$, which is an element of $H_x$, i.e. a random variable. By definition, we may identify the integral with this variable $y$, and this integral will possess the essential properties of the ordinary Riemann integral (Cramér, [12]).

The application of the theory of Hilbert space to stochastic processes seems to open very interesting possibilities. Some applications to particular classes of stochastic processes will be mentioned below. Futher important results belonging to this order of ideas will be given in a work by K. Karhunen [40], which is in course of publication.

**18. Relations to ergodic theory.** There is a close connection between the theory of stochastic processes and ergodic theory. In ergodic theory, as summarized e.g. in the treatise of E. Hopf [38], we consider an arbitrary space $\Omega$, and a probability measure $\Pi$, defined for all sets $\Sigma$ belonging to the additive class $C$. We further consider a one-parameter group of one-one transformations of $\Omega$ into itself (a "flow" in $\Omega$) such that the transformation corresponding to the parameter value $t$ takes the point $\omega = \omega_0$ into $\omega_t$, while $(\omega_t)_u = \omega_{t+u}$. Let $f(\omega)$ be a given function, defined throughout $\Omega$, and such that $f(\omega_t)$ is $C$-measurable for every fixed $t$. The well-known ergodic theorems due to von Neumann, Birkhoff, Khintchine and others are then concerned with the asymptotic behaviours of mean values, which in the classical cases are of the types

$$\frac{f(\omega_0) + f(\omega_1) + \cdots + f(\omega_{n-1})}{n}$$

or

$$\frac{1}{T}\int_0^T f(\omega_t)\,dt,$$

as $n$ or $T$ tends to infinity. (In the case of the latter expression, it is necessary to introduce some additional condition implying measurability in $t$.)

Writing $x(t, \omega) = f(\omega_t)$, it is seen that to a given transformation group $\omega \to \omega_t$ and a given function $f(\omega)$, there corresponds a stochastic process in the sense of Wiener's definition (cf. 16). The space $X_0$ of this process consists of all functions $x(t)$ representable in the form $x(t) = f(\omega_t)$, when $\omega = \omega_0$ varies over $\Omega$. The corresponding probability measure $P_0$ is defined by (16.1).

Thus any of the above-mentioned ergodic theorems may be expressed as a theorem concerning "temporal" mean values of the types

$$\frac{x(0) + x(1) + \cdots + x(n-1)}{n}$$

or

$$\frac{1}{T}\int_0^T x(t)\,dt.$$

If, according to some reasonable convergence definition, we may assign a limit to either of these expressions, as $n$ or $T$ tends to infinity, this limit will be a random variable, and it is important to find conditions which imply that this variable has a constant value for "almost all" functions $x(t)$, i.e. for all $x(t)$ except at most a set of $P_0$-measure zero.

In the particular case when $x(0)$, $x(1)$, $\cdots$ are independent variables all having the same distribution, the classical ergodic theorems yield simple cases of the laws of large numbers (cf. 10). The mean ergodic theorem of von Neumann gives the weak law, while the Birkhoff-Khintchine theorem gives the strong law. Some more general results belonging to this order of ideas will be mentioned in the sequel.

It will be seen that the two theories are largely equivalent, and it seems likely that further comparative studies of the methods will be of great value to both sides.

**19. Markoff processes.** Consider now a stochastic process, defined by a probability measure $P(S)$ in the space $X$ of all real-valued functions $x(t)$ of the real variable $t$. For any $t_1 < t_2$, there is a certain conditional probability $P(x(t_2) \subset S \mid x(t_1) = a_1)$ of the relation $x(t_2) \subset S$, relative to the hypothesis that $x(t_1)$ assumes the given value $a_1$. Suppose now that this conditional probability is independent of any additional hypothesis concerning the behaviour of $x(t)$ for $t < t_1$, so that we have e.g. for any $t_0 < t_1 < t_2$ and for any $a_0$

$$P(x(t_2) \subset S \mid x(t_1) = a_1) = P(x(t_2) \subset S \mid x(t_1) = a_1, x(t_0) = a_0).$$

In this case the process is called a *Markoff process*.

The general theory of this type of processes, which forms a natural generalization of the classical concept of Markoff chains, has been studied in basic works by Kolmogoroff [53] and Feller [26], [28]. Writing

$$P(x(t) \leqq \xi \mid x(t_0) = a_0) = F(\xi;\ t,\ a_0,\ t_0),$$

where $t_0 < t$, $F$ will be the distribution function of the random variable $x(t)$, relative to the hypothesis $x(t_0) = a_0$. Then $F$ satisfies the Chapman-Kolmogoroff equation

$$F(\xi; t, a_0, t_0) = \int_{-\infty}^{\infty} F(\xi; t, \eta, t_1)\, d_\eta F(\eta; t_1, a_0, t_0), \tag{19.1}$$

which expresses that, starting from the state $x(t_0) = a_0$, the state $x(t) \leqq \xi$ must be reached by passing through some intermediate state $x(t_1) = \eta$, where $t_0 < t_1 < t$. Subject to certain general conditions, it is possible to show that any solution of this equation satisfies certain integro-differential equations, which in some important cases reduce to partial differential equations of parabolic type, and that the d.f. $F$ is uniquely determined by these equations. However, the general conditions mentioned above are in many cases difficult to apply to particular classes of processes, and it would be important to have further investigations concerning these questions.

Markoff processes (not belonging to the subclass of differential processes, which will be considered in the following paragraph) appear in several important applications, e.g. in the theory of cosmic radiation, in certain genetical problems, in the theory of insurance risk etc. In these cases, we are often concerned with the class of *purely discontinuous* Markoff processes, where the function $x(t)$ only changes its value by jumps. If, in addition, there are only a finite or enumerable set of possible values for $x(t)$, the Chapman-Kolmogoroff equation (19.1) reduces to

$$\pi_{ik}(t_0, t) = \sum_j \pi_{ij}(t_0, t_1)\pi_{jk}(t_1, t), \tag{19.2}$$

where $\pi_{ik}(t_0, t)$ denotes the "transition probability", i.e. the probability that $x(t)$ will be in the $k$th state at the time $t$, when it is known to have been in the $i$th state at the time $t_0$. In matrix form, this equation may be written

$$\Pi(t_0, t) = \Pi(t_0, t_1)\Pi(t_1, t), \tag{19.3}$$

where $\Pi$ denotes the matrix of the $\pi_{ik}$.

When only a sequence of discrete values of $t$ are considered, we have here the classical case of Markoff chains, which has received a detailed treatment in the well-known book by Fréchet [32] (cf. also Doob, [19]). The case when $t$ is a continuous variable has been treated by Feller [28], O. Lundberg [71], Arley [1], and other authors. Some of the most important problems of this branch of the subject are concerned with the existence of a unique system of solutions of (19.2) or (19.3), and with the asymptotic behaviour of the solutions for large values of $t - t_0$. Though important results have been reached, there still remains much to be done here, and the same thing holds a fortiori with respect to the analogous problems for general Markoff processes.

**20. Differential processes.** A particularly interesting case of a Markoff process arises when, for any $\Delta t > 0$, the increment $\Delta x(t) = x(t + \Delta t) - x(t)$ is independent of $x(\tau)$ for $\tau \leqq t$. The process is then called a *differential process*. Some of the earliest studied stochastic processes belong to this class, which contains in particular the two examples discussed above in 15. Further cases of such processes arise e.g. in the theory of radioactive disintegration and in telephone technique.

Let us suppose that $x(0)$ is identically equal to zero, and that the process is uniformly continuous in probability in every finite interval $0 \leqq t \leqq T$, i.e. that for any fixed positive $\epsilon$

$$P(\mid x(t + \Delta t) - x(t) \mid > \epsilon) \rightarrow 0$$

as $\Delta t \rightarrow 0$, uniformly for $0 \leqq t \leqq T$. Then it follows from the works of Lévy, [60], [63], Khintchine [47] and Kolmogoroff [54] that, for any $t > 0$, the random variable $x(t)$ has an infinitely divisible distribution, with a characteristic function $\varphi(z; t)$ given by (9.2), where $\beta$, $\gamma$, $M(u)$ and $N(u)$ may depend on $t$.

In the particularly important case when the distribution of the increment $x(t + \Delta t) = x(t)$ does not involve $t$, but only depends on the length $\Delta t$ of the interval, we say that the process is *temporally homogeneous*, and in this case we have

$$\log \varphi(z; t) = t \log \varphi(z; 1),$$

so that we obtain the general formula for $\varphi(z; t)$ simply by replacing in (9.2) $\beta$, $\gamma$, $M(u)$ and $N(u)$ by $t\beta$, $t\gamma$, $tM(u)$ and $tN(u)$ respectively.

When $t \rightarrow \infty$, or $t \rightarrow 0$, the appropriately normalized distribution of $x(t)$ tends, under certain conditions, to a stable distribution (Cramér [7], Gnedenko [36]). When this limiting distribution is normal, there are sometimes even asymptotic expansions analogous to (11.3). Still, the problem of the asymptotic behaviour of the distribution for large $t$ does not seem to be definitely cleared up.

Khintchine [41] and Gnedenko [37] have given interesting generalizations of the law of the iterated logarithm (cf. 12) to processes of the type considered here.

The continuous process discussed in 15 in connection with the Brownian movement corresponds to the temporally homogeneous case when $\beta$, $M(u)$ and $N(u)$ all reduce to zero, so that

$$\varphi(z) = e^{-\gamma t z^2},$$

which shows that the distribution of $x(t)$ is normal, with mean zero and variance $2\gamma t$.

On the other hand, in the applications to the theory of insurance risk, $\gamma$ is zero, while $M(u)$ and $N(u)$ are connected with the distribution of the various magnitudes of claims. In this type of applications, it is often very important to find the probability that $x(t)$ satisfies an inequality of the form

$$x(t) < a + bt$$

for all values of $t$. It follows from the discussion in 16 that the definition of a probability of this type is somewhat delicate. The problem, which can be regarded as an extended form of the classical problem of "the gambler's ruin," has been solved in certain particular cases. It leads to integral equations, which in the simplest case are of the Volterra, in other cases of the Wiener-Hopf type (Cramér [6], [13], Segerdahl [79], Täcklind [81]).

**21. Orthogonal processes.** Consider now the case of a complex-valued $x(t)$, and suppose that $E \mid x(t) \mid^2$ is finite for all $t$. Without restricting the generality, we may assume that $Ex(t) = 0$ for all $t$.

Suppose now that instead of requiring, as in the case of a differential process, that the variables $x(\tau)$ and $\Delta x(t)$ should be *independent* when $\tau \leqq t$, we only lay down the less stringent condition that these variables should be *non-correlated*, i.e. that

$$E(x(\tau)\overline{\Delta x(t)}) = 0.$$

We then obtain a process which is no longer necessarily of the Markoff type. The condition implies that, for any two disjoint intervals $(t_1, t_2)$ and $(t_3, t_4)$, we have

$$E[(x(t_2) - x(t_1))(\overline{x(t_4)} - \overline{x(t_3)})] = 0,$$

so that the "chords" corresponding to two disjoint "arcs" of the curve in Hilbert space representing the process are always orthogonal (Kolmogoroff [56], [57]). A process of this type may accordingly be called an *orthogonal process.*

For a process of this type we have, writing $E \mid x(t) \mid^2 = F(t)$, $F(t + \Delta t) - F(t) = E \mid x(t + \Delta t) - x(t) \mid^2$, so that $F(t)$ is a never decreasing function of $t$. If $F(t)$ is bounded for all $t$, we shall say that the orthogonal process is bounded. For a bounded orthogonal process, the Stieltjes integral

$$\int_{-\infty}^{\infty} g(t)\, dx(t),$$

where $g(t)$ is bounded and continuous, may be defined as the limit in the mean of sums of the form

$$\sum_{\nu} g(t_\nu)(x(t_\nu) - x(t_{\nu-1})).$$

**22. Stationary processes.** When we are concerned with a process representing the temporal development of a system governed by laws which are invariant under a translation in time, it seems natural to assume that the joint distribution of any group of variables of the form

$$x(t_1 + \tau), \cdots, x(t_n + \tau) \tag{22.1}$$

is independent of $\tau$. A process satisfying this condition will be called a *stationary* process. If a stochastic process is defined by means of a "flow" $\omega \rightarrow \omega_t$ in a space $\Omega$ (cf. 18), the process will be stationary when and only when the corresponding flow is *measure-preserving*, i.e. if the transformation $\omega \rightarrow \omega_t$ changes any $C$-measurable set $S$ into a set $S_t$ of the same measure.

Under appropriate conditions with respect to the measurability of $x(t)$, the Birkhoff-Khintchine ergodic theorem holds for a stationary process, i.e. there exists a random variable $y$ such that we have

$$P_0\left(\lim_{T\to\infty} \frac{1}{T}\int_0^T x(t)\,dt = y\right) = 1, \tag{22.2}$$

where $P_0$ is the probability measure in a suitably restricted space in the sense of Doob. Further work seems to be required here, in order to make the situation quite clear, also with regard to metric transitivity.

For a stationary process, any finite moment of the joint distribution of the variables (22.1) is obviously independent of $\tau$. Suppose now that we only require that this invariance under translations in time should hold for moments of the first and second order of the joint distributions, which are assumed to be finite. The wider class of processes obtained in this way may be called *stationary of the second order*. Processes of this type have been studied for the first time by Khintchine [42]. We shall assume that $x(t)$ is complex-valued. Without restricting the generality, we may further assume that $Ex(t) = 0$ for all $t$. The product moment $E(x(t)\overline{x(u)})$ will then be a function of the difference $t - u$:

$$E(x(t)\overline{x(u)}) = R(t - u). \tag{22.3}$$

Assuming, in addition, that $R(t)$ is continuous at $t = 0$, it follows that $R(t)$ is continuous for all $t$, and the process is continuous in the sense of 17. It was shown by Khintchine that a $NS$ condition that a given function $R(t)$ should be associated with a second order stationary and continuous process by means of the relation (22.3) is that we should have

$$R(t) = \int_{-\infty}^{\infty} e^{itx}\,dF(x) \tag{22.4}$$

for all $t$, where the spectral function $F(x)$ is real, never decreasing and bounded. In particular, we have

$$F(+\infty) - F(-\infty) = R(0) = E \mid x(t) \mid^2 = \sigma^2.$$

Khintchine's condition for $R(t)$ was generalized by Cramér to the case of an arbitrary number of processes $x_1(t), \cdots, x_n(t)$, such that the product moments $E(x_i(t)\overline{x_j(u)})$ are functions of the difference $t - u$. The corresponding spectral functions $F_{ij}(x)$ are in general complex-valued and of bounded variation. Further, the expression (Cramér, [12])

$$\sum_{i,j=1}^{n} z_i \bar{z}_j \Delta F_{ij},$$

where $\Delta F_{ij} = F_{ij}(b) - F_{ij}(a)$ is, for any $a < b$, a non-negative Hermite form in the variables $z_i$. This result is closely connected with a theorem on Hilbert space considered by Kolmogoroff and Julia. It is further shown that, to any given functions $F_{ij}(x)$, $(i, j = 1, \cdots, n)$, satisfying these conditions, we can always find $n$ processes $x_1(t), \cdots, x_n(t)$ such that the joint distribution of any set of variables $x_i(t_j)$ is always *normal*, while the covariance functions $R_{ij}(t - u) = E(x_i(t)\overline{x_j(u)})$ are given by the expression

$$R_{ij}(t) = \int_{-\infty}^{\infty} e^{itx} \, dF_{ij}(x).$$

For a process $x(t)$ which is continuous and stationary of the second order, with $Ex(t) = 0$ for all $t$, we have the mean ergodic theorem

$$\text{(22.5)} \qquad \underset{T\to\infty}{\text{l.i.m.}} \frac{1}{2T} \int_{-T}^{T} e^{-\lambda it} x(t) \, dt = y$$

for any real $\lambda$. The random variable $y$ has the mean 0 and the variance $F(\lambda + 0) - F(\lambda - 0)$, where $F$ is the spectral function appearing in (22.4). If $\lambda$ is a point of continuity for $F$, it thus follows that $y = 0$ with a probability equal to 1. On the other hand, if $\lambda$ is a discontinuity, $y$ has a positive variance. Let $\lambda_1, \lambda_2, \cdots$ be all the discontinuities of $F(x)$, and let $\sigma_1^2, \sigma_2^2, \cdots$ be the corresponding saltuses, while $y_1, y_2, \cdots$ are the limits in the mean obtained from (22.5) for $\lambda = \lambda_1, \lambda_2, \cdots$. Then two different $y_j$ are always orthogonal: $E(y_j\bar{y}_k) = 0$ for $j \neq k$, and we have

$$\text{(22.6)} \qquad x(t) = \sum_{\nu} y_\nu e^{\lambda_\nu it} + \xi(t),$$

where $E\xi(t) = 0$ and

$$E \mid \xi(t) \mid^2 = \sigma^2 - \sum_{\nu} \sigma_\nu^2.$$

If $F(x)$ is a step-function, we have $\sigma^2 = \sum_\nu \sigma_\nu^2$, and it follows that $\xi(t) = 0$ with a probability equal to 1, so that (22.6) gives a "stochastic Fourier expansion" of $x(t)$ (Slutsky, [80]).

Even when $F(x)$ is arbitrary, we can obtain a spectral representation of $x(t)$ generalizing (22.6). In fact, it can be shown (Cramér, [14]) that $x(t)$ can always be represented by a Fourier-Stieltjes integral

$$x(t) = \int_{-\infty}^{\infty} e^{itu}\, dz(u), \tag{22.7}$$

where $z(u)$ is a random function attached to a bounded orthogonal process (cf. 21), such that

$$E \mid z(u + \Delta u) - z(u) \mid^2 = F(u + \Delta u) - F(u).$$

Conversely, we have

$$z(u + \Delta u) - z(u) = -\int_{-\infty}^{\infty} \frac{e^{-it(u+\Delta u)} - e^{-itu}}{2\pi i t}\, x(t)\, dt, \tag{22.8}$$

so that there is a one-one correspondence between $x(t)$ and $\Delta z(u)$. The integrals (22.7) and (22.8) are defined as limits in the mean, as shown above in 17 and 21. These results are in close correspondence with generalized harmonic analysis for an arbitrary function, as developed by Wiener [83] and Bochner [4]. The spectral representation of a stochastic process has important applications, some of which will be considered in a forthcoming paper by Karhunen [40]. An extension of the spectral representation to a more general class of processes has been given by Loève [68].

When, in particular, the $x(t)$ process is such that the joint distribution of any group of variables $x(t_1), \cdots, x(t_n)$ is normal, it follows that any increment $\Delta z(u)$ is normally distributed. Since two uncorrelated normally distributed variables are always independent, it follows that in this case the $z(u)$ process is a differential process with normally distributed increments. Important results for this case have recently been given by Doob [22].

The properties of continuity, differentiability etc. for processes of the type here considered are still incompletely known, and further work is required. A further group of important unsolved problems are connected with an interesting decomposition theorem by Wold [84], which holds for processes with a discrete time variable. The generalization of this theorem to the continuous case does not seem to have so far been given in a final form.

## REFERENCES

[1] N. Arley, "On the theory of stochastic processes and their applications to the theory of cosmic radiation," Thesis, Copenhagen, 1943.

[2] G. M. Bawly, "Ueber eine Verallgemeinerung der Grenzwertsätze der Wahrscheinlichkeitsrechnung," *Rec. Math.* (*Mat. Sbornik*), *N. S.*, Vol. 1 (1936), pp. 917–929.

[3] S. N. Bernstein, "Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes," *Math. Ann.*, Vol. 97 (1927), pp. 1–59.

[4] S. Bochner, "Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse," *Math. Ann.*, Vol. 108 (1933), pp. 378–410.

[5] H. Cramér, "On the composition of elementary errors," *Skand. Aktuarietidskr.*, Vol. 11 (1928), pp. 13–74, 141–180.

[6] ———, "On the mathematical theory of risk." Published by the Insurance Company Skandia, Stockholm, 1930.

[7] ———, "Sur les propriétés asymptotiques d'une classe de variables aléatoires," *C. R. Acad. Sci. Paris*, Vol. 201 (1935), pp. 441–443.

[8] ———, "Ueber eine Eigenschaft der normalen Verteilungsfunktion," *Math. Zeit.*, Vol. 41 (1936), pp. 405–414.

[9] ———, *Random variables and probability distributions*, Cambridge Tracts in Math., Cambridge, 1937.

[10] ———, "Sur un nouveau théorème—limite de la théorie des probabilités," *Actualités Scientifiques*, Paris, No. 736 (1938), pp. 5-23.

[11] ———, "On the representation of a function by certain Fourier integrals," *Trans. Amer. Math. Soc.*, Vol. 46 (1939), pp. 191-201. ✓

[12] ———, "On the theory of stationary random processes," *Ann. of Math.*, Vol. 41 (1940), pp. 215-230.

[13] ———, "Deux conférences sur la theorie des probabilités," *Skand. Aktuarietidskr.*, 1941, pp. 34-69.

[14] ———, "On harmonic analysis in certain functional spaces," *Ark. Mat. Astr. Fys.*, Vol. 28B (1942), pp. 1-7.

[15] ———, *Mathematical Methods of Statistics*. Princeton Univ. Press, Princeton, 1946.

[16] W. Doeblin, "Premiers éléments d'une étude systématique de l'ensemble de puissances d'une loi de probabilités," *C. R. Acad. Sci. Paris*, Vol. 206 (1938), pp. 306–308.

[17] ———, "Sur un théorème du calcul des probabilités," *C. R. Acad. Sci. Paris*, Vol. 209 (1939), pp. 742–743.

[18] J. L. Doob, "Stochastic processes depending on a continuous parameter," *Trans. Amer. Math. Soc.*, Vol. 42 (1937), pp. 107–140.

[19] ———, "Stochastic processes with an integral-valued parameter," *Trans. Amer. Math. Soc.*, Vol. 44 (1938), pp. 87–150.

[20] ———, "Regularity properties of certain families of chance variables," *Trans. Amer. Math. Soc.*, Vol. 47 (1940), pp. 455–486.

[21] ———, "The law of large numbers for continuous stochastic processes," *Duke Math. Jour.*, Vol. 6 (1940), pp. 290–306.

[22] ———, "The elementary Gaussian processes," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 229–282.

[23] J. L. Doob and M. Ambrose, "On the two formulations of the theory of stochastic processes depending upon a continuous parameter," *Ann. of Math.*, Vol. 41 (1940), pp. 737–745.

[24] P. Erdös, "On the law of the iterated logarithm," *Ann. of Math.*, Vol. 43 (1942), pp. 419–436.

[25] W. Feller, "Ueber den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung," *Math. Zeit.*, Vol. 40 (1935), pp. 521–559.

[26] ———, "Zur Theorie der stochastischen Processe (Existenz- und Eindeutigkeitssatze)," *Math. Ann.*, Vol. 113 (1936), pp. 113–160.

[27] ———, "Ueber das Gesetz der Grossen Zahlen," *Acta. Univ. Szeged.*, Vol. 8 (1937), pp. 191–201.

[28] ———, "On the integro-differential equations of purely discontinuous Markoff processes," *Trans. Amer. Math. Soc.*, Vol. 48 (1940), pp. 488–575.

[29] ———, "Generalization of a probability limit theorem of Cramér," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 361–372.

[30] ———, "The general form of the so-called law of the iterated logarithm," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 373–402.

[31] ———, "The fundamental limit theorems in probability," *Bull. Amer. Math. Soc.*, Vol. 51 (1945), pp. 800–832.

[32] M. Fréchet, *Recherches théoriques modernes sur la théorie des probabilités*, Vol. 2, Paris, 1937.

[33] B. V. Gnedenko, "Sur les fonctions caractéristiques," *Bull. Math. Univ. Moscou*, Vol. 1, (1937), pp. 16–17.

[34] ————, "On the theory of the domains of attraction of stable laws," *Učenye Zpiski, Moskovskoga gosudaistvenogo Univaziteta*, Vol. 30 (1939), pp. 61–81.

[35] ————, "On the theory of limit theorems for sums of independent random variables," *Bull. Acad. Sci. URSS*, Vol. 30 (1939), pp. 181–232, 643–647.

[36] ————, "On locally stable probability distributions," *C. R. Acad. Sci. URSS*, Vol. 35 (1942), pp. 263–266.

[37] ————, "Investigation on the growth of homogeneous random processes," *C. R. Acad. Sci. URSS*, Vol. 36 (1942), pp. 3–41.

[38] E. Hopf, *Ergodentheorie*, Ergebnisse der Mathematik, Vol. 5, No. 2, Berlin, 1937.

[39] P. L. Hsu, "The approximate distribution of the mean and of the variance of independent variates," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 1–29.

[40] K. Karhunen, Paper on stochastic processes, to appear in the *Acta Soc. Sci. Fennicae*.

[41] A. Khintchine. *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, Vol. 2, No. 4, Berlin, 1933.

[42] ————, "Korrelationstheorie der stationären stochastischen Prozesse," *Math. Ann.*, Vol. 109 (1934), pp. 604–615.

[43] ————, "Sul dominio di attrazione della legge di Gauss," *Giorn. Ist. Ital. Attuari*, Vol. 6 (1935), pp. 378–393.

[44] ————, "Su una legge dei grandi numeri generalizzata," *Giorn. Ist. Ital. Attuari*, Vol. 7 (1936), pp. 365–377.

[45] ————, "Zur Kennzeichnung der characteristischen Funktionen," *Bull. Math. Univ. Moscou*, Vol. 1 (1937), pp. 1–31

[46] ————, "Contribution à l'arithmetique des lois de distribution," *Bull. Math. Univ. Moscou*, Vol. 1 (1937), pp. 6–17.

[47] ————, "Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze," *Rec. Math. N. S.*, Vol. 2 (1937), pp. 79–117.

[48] A. Khintchine and A. Kolmogoroff, "Ueber Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden," *Rec. Math.*, Vol. 32 (1925), pp. 668–677.

[49] A. Khintchine and P. Lévy, "Sur les lois stables," *C. R. Acad. Sci. Paris*, Vol. 202 (1936), pp. 374–376.

[50] A. Kolmogoroff, "Ueber die Summen durch den Zufall bestimmter unabhängiger Grössen," *Math. Ann.*, Vol. 99 (1928) and Vol. 102 (1929), pp. 484–489.

[51] ————, "Ueber das Gesetz des iterierten Logarithmus," *Math. Ann.*, Vol. 101 (1929), pp. 126–135.

[52] ————, "Sur la loi forte des grands nombres," *C. R. Acad. Sci. Paris*, Vol. 191 (1930), pp. 910–911.

[53] ————, "Ueber die analytischen Methoden der Wahrscheinlichkeitsrechnung," *Math. Ann.*, Vol. 104 (1931), pp. 415–458.

[54] ————, "Sulla forma generale di un processo stocastico omogeneo," *Atti. Acad. naz. Lincei, Rend.*, Vol. 15 (1932), pp. 805–828, 866–869.

[55] ————, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, Vol. 2, No. 3, Berlin, 1933.

[56] ————, "Wiener Spiralen und einige andere interessante Kurven im Hilbertschen Raum," *C. R. Acad. Sci. URSS*, Vol. 26 (1940), pp. 115–118.

[57] ————, "Kurven im Hilbertschen Raum, die gegenüber einer ein parametrigen Gruppe von Bewegungen invariant sind," *C. R. Acad. Sci. URSS*, Vol. 26 (1940), pp. 6–9.

[58] P. Lévy, *Calcul des probabilités*, Paris, 1925.

[59] ————, "Sur les séries dont les termes sont des variables éventuelles indépendantes," *Studia Math.*, Vol. 3 (1931), pp. 117–155.

[60] ———, "Sur les intégrales dont les éléments sont des variables aléatoires indépendantes," *Ann. Scuola norm. super. Pisa*, (*2*), Vol. 3 (1934), pp. 337-366.

[61] ———, "Propriétés asymptotiques des sommes de variables aléatoires indépendantes," *Ann. Scuola norm. super. Pisa* (*2*), Vol. 3 (1934), pp. 347-402.

[62] ———, "La loi forte des grands nombres pour les variables aléatoires enchainées," *Jour. Math. Pures Appl.*, Vol. 15 (1936), pp. 11-24.

[63] ———, *Théorie de l'addition des variables aléatoires*, Paris, 1937.

[64] ———, "L'arithmetique des lois de probabilités," *Jour. Math. Pures Appl.*, Vol. 17 (1938), pp. 17-39.

[65] A. M. Liapounoff, "Nouvelle forme du théorème sur la limite de la probabilité," *Mémoires Acad. Saint-Petersbourg s. 8*, Vol. 12 (1901).

[66] J. W. Lindeberg, "Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung," *Math. Zeft.*, Vol. 15 (1922), pp. 211-225.

[67] W. Loève, "Étude asymptotique des sommes de variables aléatoires liées," *Jour. Math. Pures Appl.*, Vol. (9) 24 (1945), pp. 249-318.

[68] ———, "Analyse harmonique générale d'une fonction aléatoire," *C. R. Acad. Sci. Paris*, Vol. 220 (1945), pp. 380-382.

[69] F. Lundberg, "Zur Theorie der Rückversicherung," *Verhandlungen Kongr. f. Versicherungsmathematik*, Wien, 1909.

[70] ———, *Försäkringsteknisk riskutjämning*, Stockholm, 1927.

[71] O. Lundberg, "On random processes and their application to sickness and accident statistics," Thesis, Stockholm, 1940.

[72] J. Marcinkiewicz, "Sur une propriété de la loi de Gauss," *Math. Zeit.*, Vol. 4 (1938), pp. 612-618.

[73] J. Marcinkiewicz and A. Zygmund, "Sur les fonctions indépendantes," *Fund. Math.*, Vol. 29 (1937), pp. 60-90.

[74] R. E. A. C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain*, Amer. Math. Soc. Colloquium Publ., Vol. 19, New York, 1934.

[75] D. Raikov, "On the composition of Poisson laws," *C. R. Acad. Sci. URSS*, Vol. 14 (1937), pp. 9-11.

[76] ———, "On the decomposition of Gauss and Poisson laws," *Bull. Acad. Sci. URSS, Ser. Math.*, 1938, pp. 91-120.

[77] ———, "On the composition of analytic distribution functions," *C. R. Acad. Sci. URSS*, Vol. 23 (1939), pp. 511-514.

[78] I. J. Schoenberg, "Metric spaces and positive definite functions," *Trans. Amer. Math. Soc.*, Vol. 44 (1938), pp. 522-536.

[79] C. O. Segerdahl, "On homogeneous random processes and collective risk theory," Thesis, Stockholm, 1939.

[80] E. Slutsky, "Sur les fonctions aléatoires presque périodiques et sur la décomposition des fonctions aléatoires stationaires en composantes," *Actualités Scientifiques*, No. 738, Paris, 1938, pp. 33-55.

[81] S. Täcklind, "Sur le risque de ruine dans des jeux inéquitables," *Sknad. Aktuarietidskr.*, 1942, pp. 1-42.

[82] N. Wiener, "Differential space," *Jour. Math. Phys. M. I. T.*, Vol. 2 (1923), pp. 131-172.

[83] ———, "Generalized harmonic analysis," *Acta Math.*, Vol. 55 (1930), pp. 117-258.

[84] H. Wold, "A study in the analysis of stationary time series," Thesis, Stockholm, 1938.

# THE ESTIMATION OF DISPERSION FROM DIFFERENCES[1]

By Anthony P. Morse[2] and Frank E. Grubbs

*Ballistic Research Laboratory, Aberdeen Proving Ground, Maryland*

**Summary.** The estimation of variance by use of successive differences of higher order is discussed in this paper. Heretofore, attention has been focused, in published works, on estimates of variance obtained by employing the sum of squares of deviations from the mean and also by using mean square successive differences of the first order [1], [2], [3], [9]. A concise description of the method employing differences of any order with appropriate formulae for the precision of estimates so obtained and also a practical example on the use of the technique are given in section 11. Fundamental contributions to the estimation of variance from higher order differences, a study of the efficiency of the technique and proper orientation of the subject matter in the field of mathematical statistics are given in sections 2–10 of the paper.

**1. Introduction.** It frequently happens that successive observations, made at regular intervals of time, are subject to the same standard error while the means of the populations from which they are drawn display some kind of trend. The type of trend we speak of is brought about because of the manner in which we have to take measurements or because of variations in the measuring technique itself; or, again, the trend may be characteristic of the thing we are measuring. In any event, we may desire to eliminate the trend in order to study residual effects. As an example, it is desirable in the field of ballistics to evaluate the dispersion of machine guns firing from a moving airplane.

It may also happen that it is either inexpedient or impossible to estimate the standard error of the observations by the method of least squares, for in a large number of cases the type of trend is unknown. In this event a method employing differences of an appropriate order may prove valuable. The method consists merely of arranging the data in a vertical column in the order in which the observations were taken and then forming difference columns in the usual way of order 1, 2, up to say 5 or some other number depending on the peculiarities of the problem at hand and the number of the original observations. Next, sum the squares of the numbers in each column and divide the sum of squares of the $p$th order differences by $(n - p)\binom{2p}{p}$. When $n \geq 2$ and $p \geq 1$, the numbers thus arrived at are all unbiased estimates of the population variance $\sigma^2$ for the case where all the observations have the same expected value. In section 11 at the

[1] This paper is based substantially on a Ballistic Research Laboratory Report [10] of the same subject by Morse and has been prepared for publication by Grubbs at the suggestion of R. H. Kent. The authors are grateful to J. V. Lewis and H. L. Meyer for their many and varied comments, criticisms and suggestions.

[2] Now at the University of California, Berkeley, California.

end of the paper will be found a summary of this method, formulas by which the precision of the estimate of the variance $\sigma^2$ may be determined, and an example displaying the stability of this estimate with respect to $p$.

If a strong trend is present then the method of first differences will obviously yield an estimate of variance which is fictitiously large and the temptation to pass to higher order differences may quite reasonably be yielded to. As a matter of fact, unbiased estimates may be hoped for from $p$th order differences whenever there is good reason to suppose that the $p$th derivative of the trend function is small most of the time. However, even in the case of a sinusoidal trend where all derivatives have the same magnitude one may obtain good results frcm higher differences provided there are at least seven observations in each interval of length one period (see section 5 and Table II below). In connection with trends such as the sinusoidal type, the hopelessness of getting, say, even a fifth degree polynomial to fit over an interval of, say 20 periods is rather evident. It is for the above reasons that estimation of variance from higher order differences deserves consideration.

**2. Historical comment.** A brief historical development of the interest in successive differences as a means for estimating dispersion is given in [3]. This paper discusses the statistic

$$s^1 = \sqrt{\frac{\sum_{i=1}^{n/2} (x_{1i} - x_{2i-1})^2}{n}}$$

suggested by "Student" [W. S. Gossett] and E. S. Pearson and points out the relevant work of Jordan, Helmert, Vallier, Cranz, and Becker. It seems that Jordan devised methods based on sums of powers of the differences, whereas Helmert gave more careful consideration to the case of the first power, i.e. the sum of absolute differences. Reference [3] points out, however, that in these two cases all the $n(n - 1)/2$ differences that can be established frcm a sample of $n$ observations were included in the estimates of dispersion recommended by Jordan and Helmert, so that the estimate was of no value in reducing the effect of a trend. Continuing the remarks of [3], we learn that in ballistics Vallier appears to have been the first to estimate dispersion from successive differences and that Cranz and Becker commended the mean successive difference

$$E_d = \frac{\sum_{i=1}^{n-1} | x_{i+1} - x_i |}{n - 1}$$

in estimating dispersion in range of guns since they were aware of variable external effects (such as tail winds) on a projectile. In this country, Bennett [1] appears to have suggested the use of successive differences independently of European ballisticians. In this connection, Bennett suggested that the probable

TABLE 1

*The Efficiency, $W(n, p)$, of $\delta^2_{n,p}$ As An Estimate of $\sigma^2$*

| n \ p | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1.00000 | | | | | | | | | |
| 3 | .80000 | .50000 | | | | | | | | |
| 4 | .75000 | .46154 | .33333 | | | | | | | |
| 5 | .72727 | .46552 | .32000 | .25000 | | | | | | |
| 6 | .71429 | .47213 | .33149 | .24427 | .20000 | | | | | |
| 7 | .70588 | .47771 | .34453 | .25510 | .19672 | .16667 | | | | |
| 8 | .70000 | .48214 | .35537 | .26871 | .20633 | .16471 | .14286 | | | |
| 9 | .69565 | .48568 | .36408 | .28071 | .21888 | .17274 | .14159 | .12500 | | |
| 10 | .69231 | .48855 | .37113 | .29071 | .23058 | .18385 | .14830 | .12414 | .11111 | |
| 11 | .68966 | .49091 | .37691 | .29904 | .24070 | .19476 | .15802 | .12978 | .11050 | .10000 |
| 12 | .68750 | .49288 | .38173 | .30602 | .24934 | .20450 | .16798 | .13827 | .11529 | .09955 |
| 13 | .68571 | .49455 | .38580 | .31194 | .25672 | .21300 | .17714 | .14729 | .12271 | .10366 |
| 14 | .68421 | .49598 | .38928 | .31701 | .26308 | .22039 | .18530 | .15581 | .13086 | .11018 |
| 15 | .68293 | .49722 | .39228 | .32139 | .26859 | .22684 | .19250 | .16353 | .13874 | .11754 |
| 16 | .68182 | .49831 | .39490 | .32522 | .27342 | .23251 | .19887 | .17045 | .14601 | .12481 |
| 17 | .68085 | .49926 | .39721 | .32859 | .27767 | .23752 | .20452 | .17664 | .15260 | .13162 |
| 18 | .68000 | .50011 | .39925 | .33158 | .28145 | .24197 | .20956 | .18218 | .15855 | .13787 |
| 19 | .67925 | .50087 | .40107 | .33424 | .28482 | .24595 | .21407 | .18715 | .16393 | .14356 |
| 20 | .67857 | .50155 | .40271 | .33663 | .28784 | .24953 | .21813 | .19164 | .16879 | .14875 |
| 21 | .67797 | .50216 | .40419 | .33880 | .29058 | .25276 | .22181 | .19571 | .17321 | .15347 |
| 22 | .67742 | .50272 | .40553 | .34075 | .29306 | .25569 | .22515 | .19941 | .17723 | .15778 |
| 23 | .67692 | .50323 | .40675 | .34254 | .29532 | .25837 | .22819 | .20279 | .18091 | .16173 |
| 24 | .67647 | .50370 | .40787 | .34417 | .29739 | .26082 | .23098 | .20588 | .18428 | .16535 |
| 25 | .67606 | .50413 | .40889 | .34567 | .29929 | .26307 | .23354 | .20873 | .18738 | .16869 |
| 26 | .67568 | .50452 | .40984 | .34706 | .30104 | .26514 | .23590 | .21135 | .19024 | .17177 |
| 27 | .67533 | .50489 | .41071 | .34833 | .30266 | .26705 | .23809 | .21378 | .19289 | .17463 |
| 28 | .67500 | .50523 | .41152 | .34951 | .30416 | .26884 | .24012 | .21603 | .19535 | .17728 |
| 29 | .67470 | .50555 | .41228 | .35062 | .30555 | .27049 | .24200 | .21812 | .19764 | .17975 |
| 30 | .67442 | .50585 | .41298 | .35165 | .30686 | .27203 | .24375 | .22007 | .19978 | .18205 |
| 31 | .67416 | .50612 | .41363 | .35260 | .30807 | .27347 | .24539 | .22190 | .20177 | .18420 |
| 32 | .67391 | .50638 | .41425 | .35350 | .30921 | .27482 | .24693 | .22361 | .20364 | .18622 |
| 33 | .67368 | .50662 | .41482 | .35434 | .31027 | .27608 | .24837 | .22521 | .20539 | .18811 |
| 34 | .67347 | .50685 | .41536 | .35513 | .31128 | .27727 | .24973 | .22672 | .20704 | .18989 |
| 35 | .67327 | .50707 | .41587 | .35588 | .31222 | .27839 | .25101 | .22814 | .20859 | .19157 |
| 36 | .67308 | .50727 | .41635 | .35658 | .31312 | .27945 | .25221 | .22949 | .21006 | .19315 |
| 37 | .67290 | .50746 | .41671 | .35724 | .31396 | .28045 | .25335 | .23075 | .21145 | .19465 |
| 38 | .67273 | .50764 | .41724 | .35787 | .31476 | .28140 | .25443 | .23195 | .21276 | .19606 |
| 39 | .67257 | .50781 | .41765 | .35847 | .31551 | .28229 | .25545 | .23309 | .21401 | .19741 |
| 40 | .67241 | .50797 | .41804 | .35904 | .31623 | .28314 | .25642 | .23417 | .21519 | .19868 |

TABLE I—*Continued*

| p<br>n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 42 | .67213 | .50828 | .41875 | .36009 | .31756 | .28472 | .25822 | .23617 | .21738 | .20105 |
| 44 | .67188 | .50855 | .41941 | .36104 | .31877 | .28615 | .25986 | .23799 | .21937 | .20320 |
| 46 | .67164 | .50880 | .42000 | .36191 | .31987 | .28745 | .26135 | .23965 | .22118 | .20516 |
| 48 | .67143 | .50903 | .42055 | .36271 | .32088 | .28865 | .26271 | .24117 | .22284 | .20695 |
| 50 | .67123 | .50925 | .42105 | .36343 | .32180 | .28975 | .26397 | .24256 | .22437 | .20860 |
| 52 | .67105 | .50944 | .42151 | .36411 | .32266 | .29076 | .26512 | .24385 | .22578 | .21012 |
| 54 | .67089 | .50962 | .42193 | .36473 | .32345 | .29170 | .26619 | .24504 | .22708 | .21153 |
| 56 | .67073 | .50979 | .42233 | .36531 | .32418 | .29257 | .26718 | .24614 | .22829 | .21284 |
| 58 | .67059 | .50995 | .42270 | .36585 | .32487 | .29338 | .26811 | .24717 | .22941 | .21405 |
| 62 | .67033 | .51022 | .42337 | .36682 | .32609 | .29484 | .26977 | .24903 | .23144 | .21624 |
| 66 | .67010 | .51048 | .42395 | .36767 | .32718 | .29612 | .27123 | .25065 | .23322 | .21817 |
| 70 | .66990 | .51069 | .42447 | .36843 | .32813 | .29725 | .27252 | .25209 | .23479 | .21987 |
| 74 | .66972 | .51089 | .42492 | .36910 | .32898 | .29826 | .27368 | .25237 | .23619 | .22138 |
| 78 | .66957 | .51107 | .42534 | .36970 | .32975 | .29917 | .27471 | .25452 | .23745 | .22274 |
| 82 | .66942 | .51122 | .42571 | .37024 | .33043 | .29998 | .27564 | .25556 | .23859 | .22397 |
| 90 | .66917 | .51150 | .42636 | .37118 | .33162 | .30139 | .27725 | .25735 | .24055 | .22609 |
| 98 | .66897 | .51172 | .42689 | .37197 | .33262 | .30257 | .27860 | .25885 | .24219 | .22786 |
| 106 | .66879 | .51192 | .42735 | .37263 | .33346 | .30357 | .27974 | .26012 | .24358 | .22936 |
| 114 | .66864 | .51208 | .42774 | .37321 | .33418 | .30443 | .28071 | .26121 | .24477 | .23065 |
| 122 | .66851 | .51223 | .42808 | .37370 | .33482 | .30518 | .28156 | .26216 | .24581 | .23177 |
| 138 | .66829 | .51247 | .42864 | .37452 | .33585 | .30641 | .28297 | .26372 | .24752 | .23362 |
| 154 | .66812 | .51266 | .42909 | .37517 | .33667 | .30738 | .28408 | .26496 | .24887 | .23508 |
| 170 | .66798 | .51281 | .42944 | .37570 | .33734 | .30817 | .28498 | .26596 | .24997 | .23627 |
| 202 | .66777 | .51304 | .43000 | .37649 | .33836 | .30937 | .28635 | .26749 | .25164 | .23808 |
| 234 | .66762 | .51322 | .43040 | .37708 | .33909 | .31025 | .28735 | .26860 | .25285 | .23939 |
| 266 | .66751 | .51335 | .43070 | .37752 | .33965 | .31091 | .28810 | .26944 | .25377 | .24038 |
| 330 | .66734 | .51353 | .43112 | .37814 | .34044 | .31185 | .28917 | .27063 | .25508 | .24179 |
| 394 | .66723 | .51365 | .43141 | .37856 | .34097 | .31248 | .28990 | .27143 | .25596 | .24274 |
| 522 | .66709 | .51381 | .43178 | .37910 | .34164 | .31327 | .29081 | .27244 | .25707 | .24394 |
| 778 | .66695 | .51396 | .43215 | .37963 | .34233 | .31408 | .29173 | .27347 | .25819 | .24516 |
| 1290 | .66684 | .51409 | .43245 | .38007 | .34288 | .31474 | .29248 | .27430 | .25910 | .24613 |
| 2314 | .66676 | .51418 | .43264 | .38036 | .34325 | .31518 | .29298 | .27486 | .25971 | .24680 |
| ∞ | .66667 | .51429 | .43290 | .38073 | .34372 | .31573 | .29361 | .27556 | .26048 | .24763 |

error should be estimated from the root mean square successive differences as follows:

$$P.E. = .6745 \sqrt{\frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}{2(n - 1)}}.$$

In 1940, J. von Neumann and R. H. Kent in [2] investigated further the estimation of probable error from mean square successive differences (sums of squares of first differences). J. von Neumann, R. H. Kent, H. R. Bellinson, and B. I. Hart [3] considered the distribution of

$$\delta^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2$$

in a paper which appeared in June 1941. J. D. Williams [4] obtained the moments of $\eta = \frac{\delta^2}{s^2}$, where

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

and indicated that the $r$th moment of $\eta$ is equal to the $r$th moment of $\delta^2$ divided by the $r$th moment of $s^2$. The distribution of the ratio of the mean square successive difference to the variance has been published by J. von Neumann [5], [6] and B. I. Hart tabulated the probability integral and obtained percentage points for this statistic ([7], [8]). Indeed, it should be remarked that the statistical theory of successive differences is allied with the problem of serial correlation [9]. Finally, the use of squared differences of higher order than the first for estimating variance appears to have been suggested by A. A. Bennett. Quite independently, a treatment of the subject was given by Morse [10] in connection with problems on exterior ballistics. Various results on successive-difference estimation including significance tests have been given by Tintner [13]. One of Tintner's tests involves the use of selected sets of differences.

**3. Definitions and notations.** Suppose the observations $x_1, x_2, x_3, \cdots x_n$ are made at times $a = t_1 < t_2 < t_2 < \cdots < t_n = b$ and the $t_i$ are uniformly spaced without error. Let $f(t_i)$ be the true trend so that $\eta_i = f(t_i)$ is the mean of the population from which $x_i$ is drawn and $\epsilon_i = x_i - \eta_i$ is a random error. Further, let $p$ be a non-negative integer less than $n$ and denote to the $i$th backward difference of order $p$ of $x$ by $\Delta^p x_i$, i.e.

$$\Delta^p x_i = \Delta^{p-1} x_i - \Delta^{p-1} x_{i-1} = \sum_{r=0}^{p} (-1)^r \binom{p}{r} x_{i-r},$$

$$\text{where } \binom{m}{n} = \frac{m!}{n!(m-n)!}; \qquad \text{and } i \geq p + 1.$$

We define the following:

$$\delta^2_{n,p} = \frac{1}{\binom{2p}{p}(n-p)} \sum_{i=p+1}^{n} (\Delta^p \epsilon_i)^2; \tag{1}$$

$$d^2_{n,p} = \frac{1}{\binom{2p}{p}(n-p)} \sum_{i=p+1}^{n} (\Delta^p x_i)^2; \tag{2}$$

$$\nu^2_{n,p} = \frac{1}{\binom{2p}{p}(n-p)} \sum_{i=p+1}^{n} (\Delta^p \eta_i)^2; \tag{3}$$

$$k_{n,p} = \frac{2}{\binom{2p}{p}(n-p)} \sum_{i=p+1}^{n} (\Delta^p \eta_i)(\Delta^p \epsilon_i). \tag{4}$$

By $E(u)$ we will mean the expected value of $u$, whereas the variance of $u$ will be denoted by

$$\text{Var}\,(u) = E\{u - E(u)\}^2.$$

Basically, we shall assume that the $\epsilon_i$ are sufficiently Gaussian and independent that

$$E(\epsilon_i) = E(\epsilon_i^3) = 0, \qquad E(\epsilon_i^2) = \sigma^2,$$
$$\mu_4 = E(\epsilon_i^4) = 3\sigma^4,$$
$$E(\epsilon_i^\alpha \epsilon_j^\beta) = E(\epsilon_i^\alpha)E(\epsilon_j^\beta),$$

whenever $i, j, \alpha$ and $\beta$ are positive integers for which

$$i \neq j, \qquad 1 \leq i \leq n, \qquad 1 \leq j \leq n.$$

**4. Expected values.** We will now determine the mean or expected values of $\delta^2_{n,p}$ and $d^2_{n,p}$.

$$E(\delta^2_{n,p}) = \frac{1}{\binom{2p}{p}(n-p)} \sum_{i=p+1}^{n} E\left\{\sum_{r=0}^{p} (-1)^r \binom{p}{r} \epsilon_{i-r}\right\}^2,$$

$$E(\delta^2_{n,p}) = \frac{1}{\binom{2p}{p}} \sum_{r=0}^{p} \binom{p}{r}^2 \sigma^2.$$

or

$$E(\delta^2_{n,p}) = \sigma^2. \tag{5}$$

(see Lemma 1.3 of section 6 below),

Continuing, we have

$$E(d^2_{n,p}) = \frac{1}{\binom{2p}{p}(n-p)} E\left\{ \sum_{i=p+1}^{n} (\Delta^p \epsilon_i + \Delta^p \eta_i)^2 \right\},$$

$$E(d^2_{n,p}) = \frac{1}{\binom{2p}{p}(n-p)} \left\{ (n-p)\binom{2p}{p} \sigma^2 + \sum_{i=p+1}^{n} (\Delta^p \eta_i)^2 \right\},$$

or

$$E(d^2_{n,p}) = \sigma^2 + \nu^2_{n,p} . \tag{6}$$

Consequently, we observe, $d^2_{n,p}$ is on the average larger than $\sigma^2$ by the quantity $\nu^2_{n,p}$. In a particular problem, therefore, we are faced with the situation of choosing that combination of $n$ and $p$ which (*i*) regulates the size of $\nu^2_{n,p}$ and (*ii*) gives the desired precision of our estimate of variance.

**5. The magnitude of $\nu^2_{n,p}$.** In order to study the size of $\nu^2_{n,p}$, we will derive for this quantity an upper bound which will indicate the applicability of the method of differences to non-polynomial as well as polynomial trends.

Now,

$$\Delta^p \eta_i = \Delta^p f(t_i) = \int_{t_{i-1}}^{t_i} \int_0^h \cdots \int_0^h f^{(p)}(y_1 - y_2 - \cdots - y_p)\, dy_p\, dy_{p-1} \cdots dy_1 ,$$

where $t_r - t_{r-1} = h$, by straightforward integration. It will be convenient to change the order of integration; thus

$$\Delta^p f(t_i) = \int_0^h \cdots \int_0^h \int_{t_{i-1}}^{t_i} f^{(p)}(y_1 - y_2 - \cdots - y_p)\, dy_1\, dy_p \cdots dy_2 .$$

Since, from Schwarz's inequality it is clear that

$$\left\{ \int_\alpha^\beta g(s)\, ds \right\}^2 \leq (\beta - \alpha) \int_\alpha^\beta \{g(s)\}^2\, ds$$

whenever $\alpha$ and $\beta$ are real numbers and $g$ is integrable, we have

$$\{\Delta^p \eta_i\}^2 \leq h^p \int_0^h \cdots \int_0^h \int_{t_{i-1}}^{t_i} \{f^{(p)}(y_1 - y_2 - \cdots - y_p)\}^2\, dy_1\, dy_p \cdots dy_2 .$$

Also,

$$\sum_{i=p+1}^{n} \{\Delta^p \eta_i\}^2 \leq h^p \int_0^h \cdots \int_0^h \int_{t_p}^{t_n} \{f^{(p)}(y_1 - y_2 - \cdots - y_p)\}^2\, dy_1\, dy_p \cdots dy_2 .$$

But for $0 \leq r \leq (p-1)h = t_p - a$ we have

$$\int_{t_p}^{t_n} \{f^{(p)}(y_1 - r)\}^2\, dy_1 = \int_{t_p - r}^{t_n - r} \{f^{(p)}(s)\}^2\, ds \leq \int_a^b \{f^{(p)}(s)\}^2\, ds.$$

Consequently

$$\sum_{i=p+1}^{n} (\Delta^p \eta_i)^2 \leq h^p \int_0^h \cdots \int_0^h \int_a^{b^*} \{f^{(p)}(s)\}^2 \, ds \, dy_p \cdots dy_2 = h^{2p-1} \int_a^b \{f^{(p)}(s)\}^2 \, ds.$$

Since $h = \dfrac{b-a}{n-1}$, we have finally

$$\nu^2_{n,p} \leq \frac{1}{\binom{2p}{p}} \left(\frac{b-a}{n-p}\right) \left(\frac{b-a}{n-1}\right)^{2p-1} \int_a^b \frac{\{f^{(p)}(s)\}^2 \, ds}{b-a}, \tag{7}$$

which is an upper bound for $\nu^2_{n,p}$ in terms of the average value of the square of the $p$th derivative of the trend function $f$.

If the trend function $f$ is of the polynomial form,

$$f(t) = \sum_{r=0}^{p} a_r t^r$$

then the effect of the trend can be eliminated from our observations by estimating dispersion from $(p+1)$st differences. However, if it is *known* that the trend is of polynomial form, then an estimate of dispersion based on least squares would, of course, be better. In fact, it will be shown later that the precision of $\delta^2_{n,p}$ decreases markedly as $p$ increases. The use of $d^2_{n,p}$ as an estimate of $\sigma^2$ is primarily of value when the type of trend is unknown; however, even when the type of trend is known the computational simplicity of $d^2_{n,p}$ may offset to some extent its lack of optimum precision.

Let us reflect on the magnitude of $\nu^2_{n,p}$ over a single period of a sinusoidal trend, say $f(t) = \sin t$. In (7) we set $a = 0$, $b = 2\pi$ and secure

$$\nu^2_{n,p} \leq \frac{\pi}{\binom{2p}{p}(n-p)} \left(\frac{2\pi}{n-1}\right)^{2p-1}.$$

Taking $n$ to be the number of observations for a complete period, a tabulation of the upper bound for $\nu^2_{n,p}$ for this case is given in Table II. Thus, when there are about seven or more observations in each interval of length one period, estimation of dispersion from higher order differences may prove of considerable value even for this rather extreme type of trend.

**6. Some combinatorial relations.** Although we will ultimately establish expressions for the variances of $\delta^2_{n,p}$ and $d^2_{n,p}$, it appears desirable to give first a number of combinatorial relations which present themselves in the computation of moments. The relations are easily checked and most of them are possibly well known. Nevertheless, it will be convenient to record them for reference and in some instances to give proofs. In what follows it will be understood that $\binom{p}{q} = 0$ whenever $p$ and $q$ are not such integers that $0 \leq q \leq p$.

TABLE II

| $n$ / $p$ | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|
| 1 | .617 | .395 | .274 | .201 | .154 | .110 |
| 2 | .676 | .260 | .120 | .063 | .036 | .016 |
| 3 | .751 | .164 | .049 | .018 | .008 | .002 |
| 4 | .106 | .111 | .021 | .005 | .002 | .0003 |
| 5 | — | .098 | .009 | .002 | .0004 | .0000 |

Lemma 1.1. $q\binom{p}{q} = p\binom{p-1}{q-1}$.

Lemma 1.2. $\binom{p}{r} = \binom{p}{p-r}$.

Lemma 1.3. $\sum_r \binom{p}{r}\binom{p}{r+s} = \binom{2p}{p+s}$.

Proof:

$$\sum_s \binom{2p}{s} x^s = (1+x)^{2p} = \{(1+x)^p\}^2 = \left\{\sum_s \binom{p}{s} x^s\right\}^2$$

$$= \sum_s \sum_r \binom{p}{r}\binom{p}{s-r} x^s.$$

Hence

$$\binom{2p}{s} = \sum_r \binom{p}{r}\binom{p}{s-r},$$

and

$$\binom{2p}{p+s} = \sum_r \binom{p}{r}\binom{p}{p+s-r} = \sum_r \binom{p}{r}\binom{p}{r-s} = \sum_r \binom{p}{r+s}\binom{p}{r}.$$

Lemma 1.4. *If* $p^2 + r^2 > 0$ *then* $\binom{p}{r} = \binom{p-1}{r} + \binom{p-1}{r-1}$.

Lemma 1.5. $(p-2r)\binom{p}{r} = p\left\{\binom{p-1}{r} - \binom{p-1}{r-1}\right\}$.

Lemma 1.6. $(p-2r)\binom{p}{r}^2 = p\left\{\binom{p-1}{r}^2 - \binom{p-1}{r-1}^2\right\}$.

Proof: Multiply, using 1.4 and 1.5.

Lemma 1.7.[3] $r\binom{2p}{p+r}^2 = p\left\{\binom{2p-1}{p-r}^2 - \binom{2p-1}{p-r-1}^2\right\}$.

[3] Major A. A. Bennett communicated this Lemma.

PROOF: $(s - 2t)\binom{s}{t}^2 = s\left\{\binom{s-1}{t}^2 - \binom{s-1}{t-1}^2\right\}$ from 1.6.
Put $s = 2p$, $t = p - r$, then

$$2r\binom{2p}{p+r}^2 = 2p\left\{\binom{2p-1}{p-r}^2 - \binom{2p-1}{p-r-1}^2\right\}.$$

LEMMA 1.8. *If $f$ is a function, $i$, $n$, $p$ are integers and $p + 1 \leq i \leq n$, then*

$$\sum_{r=1}^{n}\binom{p}{i-r}f(i-r) = \sum_{r=0}^{p}\binom{p}{r}f(r).$$

PROOF:

$$\sum_{r=1}^{n}\binom{p}{i-r}f(i-r) = \sum_{s=i-n}^{i-1}\binom{p}{s}f(s) = \sum_{r=0}^{p}\binom{p}{r}f(r).$$

LEMMA 1.9. *If $-\infty < A(r, s) = A(s, r) < \infty$ for each integer $r$ and $s$, then*

$$E\left(\left\{\sum_{r=1}^{n}\sum_{s=1}^{n}A(r, s)\epsilon_r\epsilon_s\right\}^2\right) = (\mu_4 - 3\sigma^4)\sum_{r=1}^{n}A(r, r)^2 + \sigma^4\left\{\sum_{r=1}^{n}A(r, r)\right\}^2 + 2\sigma^4\sum_{r=1}^{n}\sum_{s=1}^{n}A(r, s)^2.$$

PROOF: Let $N(r, s) = 1$ when $r < s$ and let $N(r, s) = 0$ otherwise. Clearly

$$\sum_{r=1}^{n}\sum_{s=1}^{n}A(r, s)\epsilon_r\epsilon_s = \sum_{r=1}^{n}A(r, r)\epsilon_r^2 + 2\sum_{r=1}^{n}\sum_{s=1}^{n}N(r, s)A(r, s)\epsilon_r\epsilon_s,$$

and

$$E\left(\left\{\sum_{r=1}^{n}\sum_{s=1}^{n}A(r, s)\epsilon_r\epsilon_s\right\}^2\right) = E\left(\left\{\sum_{r=1}^{n}A(r, r)\epsilon_r^2\right\}^2\right) + 4E\left(\left\{\sum_{r=1}^{n}\sum_{s=1}^{n}N(r, s)A(r, s)\epsilon_r\epsilon_s\right\}^2\right).$$

Now

$$E\left(\left\{\sum_{r=1}^{n}A(r, r)\epsilon_r^2\right\}^2\right) = (\mu_4 - \sigma^4)\sum_{r=1}^{n}A(r, r)^2 + \sigma^4\left\{\sum_{r=1}^{n}A(r, r)\right\}^2,$$

and

$$4E\left(\left\{\sum_{r=1}^{n}\sum_{s=1}^{n}N(r, s)A(r, s)\epsilon_r\epsilon_s\right\}^2\right)$$

$$= 4\sigma^4\sum_{r=1}^{n}\sum_{s=1}^{n}N(r, s)A(r, s)^2$$

$$= 2\sigma^4\sum_{r=1}^{n}\sum_{s=1}^{n}A(r, s)^2 - 2\sigma^4\sum_{r=1}^{n}A(r, r)^2$$

The last three relations combine to yield the desired result.

Lemma 1.10. $\binom{2p}{p}^2 (n-p)^2 E(\delta^4_{n,p})$

$$= (\mu_4 - 3\sigma^4)\sum_{r=1}^{n}\left\{\sum_{i=p+1}^{n}\binom{p}{i-r}^2\right\}^2 + \sigma^4\left\{\sum_{r=1}^{n}\sum_{i=p+1}^{n}\binom{p}{i-r}^2\right\}^2$$

$$+ 2\sigma^4\sum_{r=1}^{n}\sum_{s=1}^{n}\left\{\sum_{i=p+1}^{n}\binom{p}{i-r}\binom{p}{i-s}\right\}^2.$$

Proof: Helped by 1.8, check that

$$(\Delta_i^p\epsilon)^2 = \left\{\sum_{r=0}^{p}(-1)^r\binom{p}{r}\epsilon_{i-r}\right\}^2 = \left\{\sum_{r=1}^{n}(-1)^{i-r}\binom{p}{i-r}\epsilon_r\right\}^2$$

$$= \sum_{r=1}^{n}\sum_{s=1}^{n}(-1)^{r+s}\binom{p}{i-r}\binom{p}{i-s}\epsilon_r\epsilon_s.$$

Therefore

$$\binom{2p}{p}(n-p)\delta^2_{n,p} = \sum_{r=1}^{n}\sum_{s=1}^{n}\left\{(-1)^{r+s}\sum_{i=p+1}^{n}\binom{p}{i-r}\binom{p}{i-s}\right\}\epsilon_r\epsilon_s.$$

Let

$$A(r, s) = (-1)^{r+s}\sum_{i=p+1}^{n}\binom{p}{i-r}\binom{p}{i-s},$$

and apply 1.9 to complete the proof.

Lemma 1.11.

$$\sum_{r=1}^{n}\sum_{s=1}^{n}\left\{\sum_{i=p+1}^{n}\binom{p}{i-r}\binom{p}{i-s}\right\}^2$$

$$= (n-p)\sum_{r=p-n}^{n-p}\binom{2p}{p+r}^2 - 2p\binom{2p-1}{p}^2 + 2p\binom{2p-1}{n}^2.$$

Proof.

$$\sum_{r=1}^{n}\sum_{s=1}^{n}\left\{\sum_{i=p+1}^{n}\binom{p}{i-r}\binom{p}{i-s}\right\}^2$$

$$= \sum_{i=p+1}^{n}\sum_{j=p+1}^{n}\sum_{r=1}^{n}\sum_{s=1}^{n}\binom{p}{i-s}\binom{p}{j-s}\binom{p}{i-r}\binom{p}{j-r}$$

$$= \sum_{i=p+1}^{n}\sum_{j=p+1}^{n}\sum_{r=1}^{n}\sum_{s=0}^{p}\binom{p}{s}\binom{p}{s+j-i}\binom{p}{i-r}\binom{p}{j-r}, \text{ using 1.8;}$$

$$= \sum_{i=p+1}^{n}\sum_{j=p+1}^{n}\sum_{r=0}^{p}\sum_{s=0}^{p}\binom{p}{s}\binom{p}{s+j-i}\binom{p}{r}\binom{p}{r+j-i}, \text{ using 1.8 again;}$$

$$= \sum_{i=p+1}^{n} \sum_{j=p+1}^{n} \sum_{r} \binom{p}{r}\binom{p}{r+j-i} \sum_{s} \binom{p}{s}\binom{p}{s+j-i}$$

$$= \sum_{i=p+1}^{n} \sum_{j=p+1}^{n} \left\{\sum_{r} \binom{p}{r}\binom{p}{r+j-i}\right\}^2 = \sum_{i=p+1}^{n} \sum_{j=p+1}^{n} \binom{2p}{p+j-i}^2, \text{ from 1.3;}$$

$$= \sum_{r=p+1-n}^{n-p-1} (n-p-|r|)\binom{2p}{p+r}^2$$

$$= \sum_{r=p-n}^{n-p} (n-p-|r|)\binom{2p}{p+r}^2$$

$$= (n-p)\sum_{r=p-n}^{n-p} \binom{2p}{p+r}^2 - 2\sum_{r=0}^{n-p} r\binom{2p}{p+r}^2$$

$$= (n-p)\sum_{r=p-n}^{n-p} \binom{2p}{p+r}^2 - 2\sum_{r=0}^{n-p} p\left\{\binom{2p-1}{p-r}^2 - \binom{2p-1}{p-r-1}^2\right\}, \text{ using 1.7;}$$

$$= (n-p)\sum_{r=p-n}^{n-p} \binom{2p}{p+r}^2 - 2p\left\{\binom{2p-1}{p}^2 - \binom{2p-1}{2p-n-1}^2\right\}$$

$$= (n-p)\sum_{r=p-n}^{n-p} \binom{2p}{p+r}^2 - 2p\binom{2p-1}{p}^2 + 2p\binom{2p-1}{n}^2.$$

Lemma 1.12.

$$\sum_{r=1}^{n} \sum_{i=p+1}^{n} \binom{p}{i-r}^2 = (n-p)\binom{2p}{p}.$$

Proof.

$$\sum_{r=1}^{n} \sum_{i=p+1}^{n} \binom{p}{i-r}^2 = \sum_{i=p+1}^{n} \sum_{r=1}^{n} \binom{p}{i-r}^2 = \sum_{i=p+1}^{n} \sum_{r=0}^{p} \binom{p}{r}^2, \text{ from 1.8;}$$

$$= (n-p)\sum_{r} \binom{p}{r}^2 = (n-p)\binom{2p}{p}.$$

**7. The variances of $\delta^2_{n,p}$ and $d^2_{n,p}$.** In order to get some idea as to the efficiency of the statistics $\delta^2_{n,p}$ and $d^2_{n,p}$, we will examine their variances. We have

$$\binom{2p}{p}^2 (n-p)^2 \operatorname{Var}(\delta^2_{n,p}) = \binom{2p}{p}^2 (n-p)^2 \{E(\delta^4_{n,p}) - [E(\delta^2_{n,p})]^2\}$$

$$= \binom{2p}{p}^2 (n-p)^2\sigma^4 + 2(n-p)\sigma^4 \sum_{r=p-n}^{n-p} \binom{2p}{p+r}^2 - 4p\sigma^4\binom{2p-1}{p}^2$$

$$+ 4p\sigma^4\binom{2p-1}{n}^2$$

with the aid of Lemmas 1.10, 1.11, 1.12 and using the relation $\mu_4 - 3\sigma^4 = 0$.
Thus,

$$\begin{aligned}(8)\qquad &\binom{2p}{p}^2 (n - p)^2 \operatorname{Var}(\delta^2_{n,p}) \\ &= 2(n-p)\sigma^4 \sum_{r=p-n}^{n-p} \binom{2p}{p+r}^2 - 4p\sigma^4 \binom{2p-1}{p}^2 + 4p\sigma^4 \binom{2p-1}{n}^2.\end{aligned}$$

If $2p \leq n$, then

$$\sum_{r=p-n}^{n-p} \binom{2p}{p+r}^2 \sum_r \binom{2p}{p+r}^2 = \sum_r \binom{2p}{r} = \binom{4p}{2p}.$$

Moreover, $\binom{2p-1}{n} = 0$.

Therefore,

$$(9)\qquad \binom{2p}{p}^2 (n-p)^2 \operatorname{Var}(\delta^2_{n,p}) = 2(n-p)\binom{4p}{2p}\sigma^4 - 4p\binom{2p-1}{p}^2\sigma^4$$

when $2p \leq n$.
As for the variance of $d^2_{n,p}$, we have

$$\begin{aligned}\operatorname{Var}(d^2_{n,p}) &= E\{d^2_{n,p} - \nu^2_{n,p} - \sigma^2\}^2 = E\{\delta^2_{n,p} + k_{n,p} + \nu^2_{n,p} - \nu^2_{n,p} - \sigma^2\}^2 \\ &= E\{\delta^2_{n,p} - \sigma^2) + k_{n,p}\}^2,\end{aligned}$$

or

$$(10)\qquad \operatorname{Var}(d^2_{n,p}) = \operatorname{Var}(\delta_{n,p})^2 + E(k^2_{n,p}),$$

since $E[(\delta^2_{n,p} - \sigma^2)k_{n,p}] = 0$.
However, from Schwarz's inequality, it is guaranteed that

$$E(k^2_{n,p}) \leq 4\nu^2_{n,p}\sigma^2.$$

Thus

$$(11)\qquad \operatorname{Var} d^2_{n,p} \leq \operatorname{Var}(\delta^2_{n,p}) + 4\nu^2_{n,p}\sigma^2.$$

An upper bound has already been given for $\nu^2_{n,p}$ in section 5 above.

**8. The efficiency of $\delta^2_{n,p}$.** It is appropriate to consider the efficiency (as defined by Fisher [11]) of the statistic $\delta^2_{n,p}$. In this sense, the efficiency of $\delta^2_{n,p}$ is given by

$$W(n, p) = \frac{\operatorname{Var} s^2_n}{\operatorname{Var} \delta^2_{n,p}}. \qquad \text{where } s^2_n = \frac{\sum (x_i - \bar{x})^2}{n - 1}.$$

Accordingly,

$$W(n, p) = \frac{2\sigma^4}{(n - 1)\,\mathrm{Var}\,(\delta^2_{n,p})}$$

or

$$W(n, p) \tag{12}$$
$$= \frac{(n - p)^2 \binom{2p}{p}^2}{(n - 1)\left\{(n - p) \sum_{r=p-n}^{n-p} \binom{2p}{p + r}^2 - 2p\binom{2p - 1}{p}^2 + 2p\binom{2p - 1}{n}^2\right\}}.$$

If $2p \leq n$

$$W(n. p) = \frac{(n - p)^2 \binom{2p}{p}^2}{(n - 1)\left\{(n - p)\binom{4p}{2p} - 2p\binom{2p - 1}{p}^2\right\}}, \qquad \text{from (9);} \tag{13}$$

or

$$W(n, p) = \frac{\binom{2p}{p}^2}{\binom{4p}{2p}\left\{1 - \frac{p - 1}{n - p}\right\}\left\{1 - \frac{2p\binom{2p - 1}{p}^2}{(n - p)\binom{4p}{2p}}\right\}}, \qquad \text{if } 2p \leq n. \tag{14}$$

Formulas (12) and (13) were used in preparing Table I given at the end of the paper. For convenience in using formulas (1) and (2) the binomial coefficients $\binom{2p}{p}$ for $0 \leq p \leq 10$ are given in Table III.

If $n \geq 2$, then

$$W(n, 1) = \frac{2}{3}\cdot\frac{1}{1 - \frac{1}{3n - 3}} = \frac{2(n - 1)}{3n - 4}, \tag{15}$$

as was pointed out by von Neumann, Kent, Bellinson, and Hart in [3].

If $n \geq 4$, then

$$W(n, 2) = \frac{18}{35}\,\frac{1}{\left\{1 + \frac{1}{n - 2}\right\}\left\{1 - \frac{18}{35(n - 2)}\right\}} = \frac{18(n - 2)^2}{(n - 1)(35n - 88)}. \tag{16}$$

As a limiting value for $n$, we have

$$W(\infty, p) = \lim_{n \to \infty} W(n, p) = \frac{\binom{2p}{p}^2}{\binom{4p}{2p}}. \tag{17}$$

Using Stirling's formula for the approximation to the factorial, we have

$$\lim_{p \to \infty} \sqrt{p}\, W(\infty, p) = \sqrt{\frac{2}{\pi}}.$$

Thus, as $p \to \infty$, $W(\infty, p)$ tends to zero and is asymptotically equal to $\sqrt{\dfrac{2}{\pi p}}$

TABLE III

*The Binomial Coefficient* $\binom{2p}{p}$

| $p$ | $\binom{2p}{p}$ |
|---|---|
| 0 | 1 |
| 1 | 2 |
| 2 | 6 |
| 3 | 20 |
| 4 | 70 |
| 5 | 252 |
| 6 | 924 |
| 7 | 3432 |
| 8 | 12870 |
| 9 | 48620 |
| 10 | 184756 |

For the case $n \geq 2$, $p \geq 1$ and $f$ *constant*, then $s_n^2 = \dfrac{\Sigma\,(x_i - \bar{x})^2}{n - 1}$ and $\delta^2_{n,p}$ and $d^2_{n,p}$ are all unbiased estimates of the population variance $\sigma^2$. Moreover, for this case

$$W(n, p) = \frac{\mathrm{Var}\,(s_n^2)}{\mathrm{Var}\,(\delta^2_{n,p})} = \frac{\mathrm{Var}\,(s_n^2)}{\mathrm{Var}\,(d^2_{n,p})}.$$

Using $s_m^2$ based on $m - 1$ degrees of freedom and keeping the trend, $f$, constant, then $m$ and $n$ may be chosen so that approximately

$$\mathrm{Var}\,(s_m^2) = \mathrm{Var}\,(d^2_{n,p})$$

and for a normal population this means that

$$m = 1 + (n - 1)W(n, p).$$

Using Table I, it may be seen that for constant trend, $f$, the worth of $d^2_{50,10}$ as an estimate of $\sigma^2$ for a normal population is about the same as that of $s^2_{11}$, whereas that of $d^2_{50,1}$ is about equivalent to $s^2_{40}$. However, if the trend $f$ is not constant then the worth of $s^2_n$ as an estimate of $\sigma^2$ is diminished while that of $d^2_{n,p}$ is increased.

Similarly, if the trend is cubic over 20 observations then least squares gives an unbiased estimate of $\sigma^2$ based on 16 degrees of freedom, whereas $d^2_{20,4}$ gives an estimate equivalent in precision to about 6.4 degrees of freedom. However, if only eight observations follow a cubic trend, then least squares furnish an unbiased estimate of $\sigma^2$ based on four degrees of freedom whereas $d^2_{8,4}$ furnishes an estimate equivalent to about 1.9 degrees of freedom. Thus, in the case of 20 observations, cubic least squares is, so to speak, 2.5 times as valuable as $d^2_{20,4}$; in the case of eight observations, cubic least squares is 2.1 times as valuable as $d^2_{8,4}$.

It might be mentioned that the method of differences is of value in estimating goodness of fit. If the fit is good, then our estimate of $\sigma^2$ derived from least squares should on the average be equal to the estimate derived from a suitable $d^2_{n,p}$. If the fit is poor then $d^2_{n,p}$ will be smaller on the average than the former.

**9. The approximate probable error in estimating $\sigma$ from differences.** The approximate standard error of $\delta_{n,p}$ is given by the relation

$$\text{S.E. } (\delta_{n,p}) \sim \frac{1}{2}\frac{\text{S.E. } (\delta^2_{n,p})}{\sigma} = \frac{\sigma}{\sqrt{2(n-1)W(n,p)}}.$$

If $p$ has been so chosen that $\nu^2_{n,p}$ is suitably small then [see equation (11)] some confidence may be put in the approximate formulas:

$$\text{S.E } (d_{n,p}) = \frac{\sigma}{\sqrt{2(n-1)W(n,p)}} \tag{18}$$

$$\text{P.E. } (d_{n,p}) = \frac{.6745\sigma}{\sqrt{2(n-1)W(n,p)}}. \tag{19}$$

Formula (19) was used in preparing Table IV which gives the approximate probable error to be feared in using $d_{n,p}$ as an estimate of $\sigma$. This table should yield interesting information whenever $p$ has been chosen so that $d^2_{n,p}$ is a suitably unbiased estimate of $\sigma^2$.

**10. Remarks.** We have presented a useful technique for estimating variance from higher order differences and have given the precision of our estimate. The method of estimating variance from higher order differences appears to be quite valuable in cases where the type of trend in our observations is unknown. A considerable field of work remains concerning a complete investigation of the distribution and other properties of the statistic $d^2_{n,p}$. In this connection, Baer [12] has already published a study on the stochastic limit of $\frac{n}{n-1} d^2_{n,1}$. It is hoped that others will contribute to the problem of estimating dispersion

TABLE IV

*The Probable Error In Estimating σ From Differences**

| p<br>n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | .4769 | | | | | | | | | | |
| 2 | .3373 | .4769 | | | | | | | | | |
| 3 | .2753 | .3771 | .4769 | | | | | | | | |
| 4 | .2384 | .3180 | .4054 | .4769 | | | | | | | |
| 5 | .2133 | .2796 | .3495 | .4215 | .4769 | | | | | | |
| 6 | .1948 | .2524 | .3104 | .3704 | .4404 | .4769 | | | | | |
| 7 | .1803 | .2317 | .2817 | .3318 | .3855 | .4390 | .4769 | | | | |
| 8 | .1686 | .2154 | .2596 | .3024 | .3477 | .3969 | .4442 | .4769 | | | |
| 9 | .1589 | .2022 | .2420 | .2794 | .3183 | .3604 | .4057 | .4481 | .4769 | | |
| 10 | .1508 | .1911 | .2274 | .2610 | .2948 | .3311 | .3708 | .4128 | .4513 | .4769 | |
| 11 | .1438 | .1816 | .2153 | .2457 | .2758 | .3074 | .3417 | .3794 | .4186 | .4537 | .4769 |
| 12 | .1376 | .1734 | .2048 | .2328 | .2599 | .2880 | .3180 | .3508 | .3867 | .4234 | .4558 |
| 13 | .1323 | .1663 | .1958 | .2217 | .2465 | .2717 | .2983 | .3272 | .3587 | .3930 | .4276 |
| 14 | .1274 | .1599 | .1878 | .2120 | .2350 | .2579 | .2818 | .3073 | .3351 | .3656 | .3984 |
| 15 | .1231 | .1542 | .1808 | .2035 | .2248 | .2459 | .2677 | .2905 | .3152 | .3423 | .3718 |
| 16 | .1192 | .1491 | .1744 | .1960 | .2159 | .2355 | .2554 | .2761 | .2983 | .3223 | .3485 |
| 17 | .1156 | .1445 | .1687 | .1892 | .2080 | .2262 | .2447 | .2637 | .2837 | .3052 | .3286 |
| 18 | .1124 | .1403 | .1636 | .1831 | .2009 | .2180 | .2352 | .2527 | .2710 | .2905 | .3116 |
| 19 | .1094 | .1364 | .1589 | .1775 | .1945 | .2106 | .2267 | .2430 | .2599 | .2777 | .2967 |
| 20 | .1066 | .1328 | .1545 | .1724 | .1886 | .2040 | .2191 | .2343 | .2500 | .2663 | .2837 |
| 21 | .1040 | .1295 | .1505 | .1677 | .1832 | .1978 | .2121 | .2264 | .2411 | .2562 | .2722 |
| 22 | .1016 | .1265 | .1468 | .1634 | .1783 | .1922 | .2058 | .2193 | .2331 | .2472 | .2620 |
| 23 | .0994 | .1236 | .1433 | .1594 | .1738 | .1871 | .2000 | .2129 | .2258 | .2391 | .2529 |
| 24 | .0973 | .1209 | .1401 | .1557 | .1695 | .1824 | .1948 | .2069 | .2191 | .2316 | .2446 |
| 25 | .0954 | .1184 | .1371 | .1522 | .1656 | .1779 | .1898 | .2015 | .2131 | .2249 | .2370 |
| 26 | .0935 | .1160 | .1343 | .1490 | .1619 | .1739 | .1853 | .1964 | .2075 | .2187 | .2301 |
| 27 | .0918 | .1138 | .1316 | .1459 | .1585 | .1700 | .1810 | .1917 | .2023 | .2130 | .2238 |
| 28 | .0902 | .1117 | .1291 | .1431 | .1553 | .1664 | .1770 | .1873 | .1975 | .2077 | .2180 |
| 29 | .0885 | .1097 | .1268 | .1404 | .1522 | .1631 | .1733 | .1832 | .1930 | .2028 | .2126 |
| 30 | .0871 | .1078 | .1245 | .1378 | .1493 | .1599 | .1698 | .1794 | .1888 | .1981 | .2076 |
| 31 | .0857 | .1060 | .1224 | .1354 | .1466 | .1569 | .1665 | .1758 | .1848 | .1938 | .2029 |
| 32 | .0843 | .1043 | .1204 | .1331 | .1441 | .1540 | .1634 | .1724 | .1811 | .1898 | .1985 |
| 33 | .0831 | .1027 | .1184 | .1309 | .1416 | .1514 | .1605 | .1692 | .1776 | .1860 | .1944 |
| 34 | .0818 | .1012 | .1166 | .1288 | .1393 | .1488 | .1577 | .1661 | .1744 | .1825 | .1905 |
| 35 | .0807 | .0999 | .1149 | .1268 | .1371 | .1464 | .1550 | .1632 | .1713 | .1791 | .1869 |
| 36 | .0795 | .0983 | .1132 | .1249 | .1350 | .1441 | .1525 | .1605 | .1683 | .1759 | .1834 |
| 37 | .0784 | .0969 | .1116 | .1231 | .1330 | .1418 | .1501 | .1579 | .1655 | .1729 | .1802 |
| 38 | .0774 | .0956 | .1101 | .1214 | .1311 | .1397 | .1478 | .1555 | .1628 | .1700 | .1771 |
| 39 | .0764 | .0943 | .1086 | .1197 | .1292 | .1377 | .1456 | .1531 | .1603 | .1673 | .1741 |
| 40 | .0754 | .0931 | .1072 | .1181 | .1274 | .1358 | .1435 | .1508 | .1578 | .1646 | .1713 |

TABLE IV—*Continued*

| n \ p | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 42 | .0736 | .0909 | .1045 | .1151 | .1241 | .1322 | .1396 | .1466 | .1533 | .1597 | .1661 |
| 44 | .0719 | .0887 | .1020 | .1123 | .1211 | .1288 | .1360 | .1427 | .1491 | .1553 | .1613 |
| 46 | .0703 | .0868 | .0997 | .1097 | .1182 | .1257 | .1326 | .1391 | .1453 | .1512 | .1570 |
| 48 | .0689 | .0849 | .0975 | .1073 | .1155 | .1228 | .1295 | .1357 | .1417 | .1474 | .1529 |
| 50 | .0675 | .0832 | .0955 | .1050 | .1130 | .1201 | .1266 | .1326 | .1383 | .1438 | .1492 |
| 52 | .0661 | .0815 | .0936 | .1029 | .1107 | .1176 | .1238 | .1297 | .1352 | .1405 | .1457 |
| 54 | .0649 | .0800 | .0918 | .1009 | .1085 | .1152 | .1213 | .1270 | .1323 | .1375 | .1425 |
| 56 | .0637 | .0785 | .0901 | .0990 | .1064 | .1129 | .1189 | .1244 | .1296 | .1346 | .1394 |
| 58 | .0626 | .0771 | .0885 | .0972 | .1045 | .1108 | .1166 | .1220 | .1271 | .1319 | .1366 |
| 62 | .0606 | .0746 | .0855 | .0939 | .1008 | .1069 | .1125 | .1176 | .1224 | .1270 | .1313 |
| 66 | .0587 | .0723 | .0828 | .0909 | .0975 | .1034 | .1087 | .1136 | .1182 | .1225 | .1266 |
| 70 | .0570 | .0702 | .0804 | .0881 | .0946 | .1002 | .1053 | .1100 | .1144 | .1185 | .1224 |
| 74 | .0554 | .0682 | .0781 | .0856 | .0919 | .0973 | .1022 | .1067 | .1109 | .1149 | .1186 |
| 78 | .0540 | .0664 | .0760 | .0833 | .0894 | .0947 | .0994 | .1037 | .1077 | .1115 | .1152 |
| 82 | .0527 | .0648 | .0741 | .0812 | .0871 | .0922 | .0968 | .1009 | .1048 | .1085 | .1120 |
| 90 | .0503 | .0618 | .0707 | .0774 | .0830 | .0878 | .0921 | .0960 | .0997 | .1031 | .1063 |
| 98 | .0482 | .0592 | .0677 | .0741 | .0794 | .0840 | .0880 | .0917 | .0952 | .0984 | .1014 |
| 106 | .0463 | .0569 | .0650 | .0712 | .0762 | .0806 | .0845 | .0880 | .0913 | .0943 | .0972 |
| 114 | .0447 | .0549 | .0627 | .0686 | .0734 | .0776 | .0813 | .0847 | .0878 | .0907 | .0934 |
| 122 | .0432 | .0530 | .0606 | .0663 | .0709 | .0749 | .0785 | .0817 | .0847 | .0875 | .0900 |
| 138 | .0406 | .0498 | .0569 | .0622 | .0666 | .0703 | .0736 | .0766 | .0794 | .0819 | .0843 |
| 154 | .0384 | .0472 | .0538 | .0589 | .0630 | .0664 | .0695 | .0723 | .0749 | .0773 | .0795 |
| 170 | .0366 | .0449 | .0512 | .0560 | .0599 | .0632 | .0661 | .0687 | .0711 | .0734 | .0755 |
| 202 | .0336 | .0412 | .0470 | .0513 | .0548 | .0578 | .0605 | .0629 | .0650 | .0671 | .0689 |
| 234 | .0312 | .0382 | .0436 | .0476 | .0509 | .0537 | .0561 | .0583 | .0603 | .0621 | .0639 |
| 266 | .0292 | .0359 | .0409 | .0446 | .0477 | .0503 | .0525 | .0546 | .0565 | .0582 | .0598 |
| 330 | .0262 | .0322 | .0367 | .0400 | .0428 | .0451 | .0471 | .0489 | .0505 | .0521 | .0535 |
| 394 | .0240 | .0295 | .0336 | .0366 | .0391 | .0412 | .0430 | .0447 | .0462 | .0475 | .0488 |
| 522 | .0209 | .0256 | .0292 | .0318 | .0339 | .0357 | .0373 | .0387 | .0400 | .0412 | .0423 |
| 778 | .0171 | .0210 | .0239 | .0260 | .0278 | .0292 | .0305 | .0317 | .0327 | .0337 | .0346 |
| 1290 | .0133 | .0163 | .0185 | .0202 | .0216 | .0227 | .0237 | .0246 | .0254 | .0261 | .0268 |
| 2314 | .0099 | .0121 | .0138 | .0151 | .0161 | .0169 | .0177 | .0183 | .0189 | .0195 | .0200 |

* If $d_{n,p}^2$ is a sufficiently unbiased estimate of $\sigma^2$, then the approximate probable error to be feared in using $d_{n,p}$ as an estimate of $\sigma$ may be obtained by multiplying the following tabular entries by $\sigma$.

when observed data display trends as it is believed that the method of differences deserves much attention. In particular, it is hoped that someone will have the time and ingenuity to calculate the distribution of the statistic

$$\frac{\delta^2_{n,p}}{\delta^2_{n,p+1}}.$$ [4]

Were this done, an admirable criterion would be at hand for gauging the significance of a change in the estimate of $\sigma^2$ as we pass from differences of order $p$ to those of order $p + 1$. Of course, useful information in this connection could be obtained from a knowledge of the distributions of $\delta^2_{n,p}$ and $\delta^2_{n,p+1}$ ; in fact their variances as herein calculated give us a basis for somewhat reasonable conclusions. An expression for the standard error of the difference between the estimates of $\sigma^2$ from two consecutive series of finite differences is given in [13, Chapter VI].

In connection with testing goodness of fit, it would be valuable also to know the distribution of

$$\frac{S^2_{n,p}}{\delta^2_{n,p+1}},$$

where $S^2_{n,p}$ is the estimate of variance derived from the least squares fitting of a polynomial of degree $p$.

For convenience of reference, we conclude the paper with

**11. A concise description of the method and its precision.** It frequently happens that successive observations made at regular intervals are subject to the same standard error $\sigma$ while the means of the populations from which they are drawn display a trend. We give here a method of estimating the variance $\sigma^2$ and of determining the precision of our estimate. This method is primarily of value when the trend is unknown; however even when the type of trend is known, its computational simplicity may make the method advantageous.

*The method.* Arrange the data in a vertical column and then in the usual way form difference columns of order 1, 2, $\cdots$ , $p$. Sum the squares of the $p$th order differences and divide by the number $(n - p)\binom{2p}{p}$. Our estimate of $\sigma^2$ is the number $d^2_{n,p}$ , where

$$d^2_{n,p} = \frac{1}{\binom{2p}{p}(n - p)} \sum_{i=p+1}^{n} (\Delta^p x_i)^2.$$

[4] Dixon [9] gives moments of the statistic $\dfrac{\sum_{i=1}^{n} (x_i - 2x_{i+1} + x_{i+2})^2}{\sum_{i=1}^{n} (x_i - x_{i+1})^2}$ where $x_{n+1} = x_1$ and $x_{n+2} = x_2$.

*The precision.* The precision of this estimate may be determined from the following information (which has been derived in the present paper):

$E(d^2_{n,p}) = \sigma^2 + \nu^2_{n,p}$ ;

$$\nu^2_{n,p} \leq \frac{1}{\binom{2p}{p}} \left(\frac{b-a}{n-p}\right)\left(\frac{b-a}{n-1}\right)^{2p-1} \int_a^b \frac{[f^{(p)}(s)]^2\, ds}{b-a} ;$$

$$\operatorname{Var}(d^2_{n,p}) \leq \operatorname{Var}(\delta^2_{n,p}) + 4\nu^2_{n,p}\sigma^2 ;$$

$$\operatorname{Var}(\delta^2_{n,p}) = \frac{2\sigma^4}{(n-1)W(n,p)},$$

where $W(n, p)$ is given in Table I.

TABLE V

| p | $\sigma_x$ | $\sigma_y$ | $\sigma_z$ |
|---|---|---|---|
| 1 | 18.90 | 184.62 | 11.22 |
| 2 | 1.21 | 1.88 | 10.56 |
| 3 | .88 | 1.85 | 10.30 |
| 4 | .87 | 1.84 | 10.12 |
| 5 | .86 | 1.83 | 10.01 |

In case $\nu^2_{n,p}$ is sufficiently small (this is determined by the requirements of the given problem), then Table IV may be used directly to determine the approximate probable error in using $d_{n,p}$ as an estimate of $\sigma$.

*An example.* As a practical example of the use of the method of differences when the trend is unknown and of the stability of the statistic $d^2_{n,p}$ with respect to $p$, we mention a recent problem at Aberdeen Proving Ground which had to do with estimating the accuracy with which certain photographic measurements locate a moving object. Ballistic Cameras were used to determine horizontal $x$ and $y$, and vertical $z$ coordinates (all in feet) of an airplane traveling about 160 mph at an elevation of about 35,000 feet. An automatic pilot was in use in the airplane as it flew over a three mile course. At one second intervals for a period of 70 seconds two Ballistic Cameras, 5000 feet apart, were used to locate the plane. Since the plane was traveling pretty much in the $y$ direction one would expect: that first differences would yield a standard error in $y$ far in excess of its true one; that second differences would furnish a much better estimate; and that perhaps third differences would yield a still more trustworthy one. No matter what order of difference is used we never expect such an estimate to be too small. In this problem, the standard errors in $x$, $y$, $z$ as estimated from differences of certain orders, $p$, were as given in Table V.

## REFERENCES

[1] A. A. Bennett, unpublished report to the Chief of Ordnance, U. S. Army, circa 1918.

[2] R. H. Kent and J. von Neumann, "The estimation of the probable error from successive differences", Ballistic Res. Lab. Report No. 175, Feb., 1940.

[3] John von Neumann, R. H. Kent, H. R. Bellinson, and B. I. Hart, "The mean square successive difference", *Annals of Math. Stat.*, Vol. 12 (1941), pp. 153-162.

[4] J. D. Williams, "Moments of the ratio of the mean square successive difference to the mean square difference in samples from a normal universe", *Annals of Math. Stat.*, Vol. 12 (1941), pp. 239–241.

[5] John von Neumann, "Distribution of the ratio of the mean square successive difference to the variance", *Annals of Math. Stat.*, Vol. 12 (1941), pp. 367–395.

[6] John von Neumann, "A further remark concerning the distribution of the ratio of the mean square successive difference to the variance", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 86–88.

[7] B. I. Hart, "Tabulation of the probabilities for the ratio of the mean square successive difference to the variance", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 207–214.

[8] B. I. Hart, "Significance levels for the ratio of the mean square successive difference to the variance", *Annals of Math. Stat.*, Vol. 13 (1942), pp. 445–446.

[9] Wilfrid J. Dixon, "Further contributions to the problem of serial correlation", *Annals of Math. Stat.*, Vol. 15 (1944), pp. 119–144.

[10] Anthony P. Morse, "The estimation of dispersion from differences", Ballistic Res. Lab. Report No. 557, July, 1945.

[11] R. A. Fisher, *Phil. Trans. A.*, Vol. 222 (1922), p. 316.

[12] Reinhold Baer, "Sampling from a changing population", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 348–361.

[13] G. Tintner, *The variate Difference Method*, (Cowles Commission Monograph No. 5), Principia Press, Bloomington, Indiana, 1940.

# THE EFFICIENCY OF SEQUENTIAL ESTIMATES AND WALD'S EQUATION FOR SEQUENTIAL PROCESSES

By J. Wolfowitz

*Columbia University*

**1. Summary.** Let $n$ successive independent observations be made on the same chance variable whose distribution function $f(x, \theta)$ depends on a single parameter $\theta$. The number $n$ is a chance variable which depends upon the outcomes of successive observations; it is precisely defined in the text below. Let $\theta^*(x_1, \cdots, x_n)$ be an estimate of $\theta$ whose bias is $b(\theta)$. Subject to certain regularity conditions stated below, it is proved that

$$\sigma^2(\theta^*) \geq \left(1 + \frac{db}{d\theta}\right)^2 \left[EnE\left(\frac{\partial \log f}{\partial \theta}\right)^2\right]^{-1}.$$

When $f(x, \theta)$ is the binomial distribution and $\theta^*$ is unbiased the lower bound given here specializes to one first announced by Girshick [3], obtained under no doubt different conditions of regularity. When the chance variable $n$ is a constant the lower bound given above is the same as that obtained in [2], page 480, under different conditions of regularity.[1]

Let the parameter $\theta$ consist of $l$ components $\theta_1, \cdots, \theta_l$ for which there are given the respective unbiased estimates $\theta_1^*(x_1, \cdots, x_n), \cdots, \theta_l^*(x_1, \cdots, x_n)$. Let $\| \lambda_{ij} \|$ be the non-singular covariance matrix of the latter, and $\| \lambda^{ij} \|$ its inverse. The concentration ellipsoid in the space of $(k_1, \cdots, k_l)$ is defined as

$$\sum_{i,j} \lambda^{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2.$$

(This valuable concept is due to Cramér). If a unit mass be uniformly distributed over the concentration ellipsoid, the matrix of its products of inertia will coincide with the covariance matrix $\| \lambda_{ij} \|$. In [4] Cramér proves that no matter what the unbiased estimates $\theta_1^*, \cdots, \theta_l^*$, (provided that certain regularity conditions are fulfilled), when $n$ is constant their concentration ellipsoïd always contains within itself the ellipsoid

$$\sum_{i,j} \mu_{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2$$

where

$$\mu_{ij} = nE\left(\frac{\partial \log f}{\partial \theta_i}\frac{\partial \log f}{\partial \theta_j}\right).$$

[1] To whom this result is to be ascribed is not clear from the context in which Professor Cramér describes it (in [2]). After the present paper was completed the author learned of the papers by Rao [8] and Aitken and Silverstone [9], both of which deal with this question. The author is indebted to Prof. M. S. Bartlett for drawing his attention to these papers.

Consider now the sequential procedure of this paper. Let $\theta_1^*, \cdots, \theta_l^*$ be, as before, unbiased estimates of $\theta_1, \cdots, \theta_l$, respectively, recalling, however, that the number $n$ of observations is a chance variable. It is proved that the concentration ellipsoid of $\theta_1^*, \cdots, \theta_l^*$ always contains within itself the ellipsoid

$$\sum_{i,j} \mu'_{ij}(k_i - \theta_i)\,(k_j - \theta_j) = l + 2$$

where

$$\mu'_{ij} = EnE\left(\frac{\partial \log f}{\partial \theta_i}\frac{\partial \log f}{\partial \theta_j}\right).$$

When $n$ is a constant this becomes Cramér's result (under different conditions of regularity).

In section 7 is presented a number of results related to the equation $EZ_n = EnEX$, which is due to Wald [6] and is fundamental for sequential analysis.

## 2. Introduction.

Let X be a chance variable whose distribution function $f(x, \theta)$ depends on the parameter $\theta$. It is assumed that $X$ either has a probability density function (which we then denote by $f(x, \theta)$) or that it can take only an at most denumerable number of discrete values (in the latter case $f(x, \theta) = P\{X = x\}$, where the latter symbol denotes the probability of the relation in braces). Let $\omega = x_1, x_2, \cdots$ be an infinite sequence of observations on $X$, and let $\Omega$ be the space of "points" $\omega$. Let there be given an infinite sequence of Borel measurable functions $\varphi_1(x_1), \varphi_2(x_1, x_2), \cdots, \varphi_j(x_1, \cdots, x_j), \cdots$ defined for all $\omega$ in $\Omega$, such that each takes only the values zero and one. It is well known that the function $f(x, \theta)$ defines a measure (probability) on a Borel field in $\Omega$. We assume that everywhere in $\Omega$, except possibly on a set whose probability is zero for all $\theta$ under consideration, at least one of the functions $\varphi_1, \varphi_2, \cdots$ takes the value one. Let $n(\omega)$ be the smallest integer at which this occurs. Thus $n(\omega)$ is a chance variable.

In statistical applications the chance variable $n(\omega)$ may be interpreted as a rule for terminating a sequence of observations on the chance variable $X$, the probability of termination being one, and the decision to terminate depending only upon the observations obtained. A sequential test is an example of this procedure. The converse is, however, not true, because the process described above does not require that any statistical decision should be reached when the process of drawing observations is terminated.

An "estimate" of $\theta$ is a function $\theta^*(x_1, \cdots, x_n)$ of the observations $x_1, \cdots, x_n$ (those obtained prior to the "termination" of the process of drawing observations). In the sequel we shall limit ourselves to estimates whose second moments are finite. The estimate is "unbiased" if $E\theta^*$, the expected value of $\theta^*$, is $\theta$. When this is not so $E\theta^* - \theta$ is called the bias, $b(\theta)$, of $\theta^*$. In general the bias is a function of $\theta$. It is obvious that the function $\theta^*$ may be undefined on a set of points $(x_1, \cdots, x_n)$ whose probability is zero for all $\theta$ under consideration.

In the present paper we shall be concerned with an upper bound on the efficiency of a sequential estimate, or, more precisely, with a lower bound on its variance. This lower bound is intimately related to certain results on the efficiency of the maximum likelihood estimate from a sample of fixed size. This is not surprising since fixed-size sampling is a special instance of sequential sampling. The results obtained in this paper are also obviously and intimately related to those due to Cramér [4] and those described by him in [2], pp. 477–488. Naturally the conditions of regularity (restrictions on $f(x, \theta)$, $\theta^*$, etc.) under which the results are proved are different. For example, no restrictions on the sequential sampling procedure need appear in the statement of a theorem which deals only with samples of fixed size.

The argument below proceeds as if $f(x, \theta)$ were a probability density function. The results apply equally well to the case where $f(x, \theta)$ is the probability function of a discrete chance variable provided:

1). Integration is replaced by summation wherever this is obviously required.

2). The phrase "almost all points" in a Euclidean space of any finite dimensionality is understood

a). as all points in the space with the possible exception of a set of Lebesgue measure zero, when $f(x, \theta)$ is a probability density function

b). as all points in the space with the possible exception of points one of whose coordinates is a member of the set $Z$, when $f(x, \theta)$ is the probability function of a discrete chance variable. The set $Z$ consists of all points $z$ such that $f(z, \theta) = 0$ identically for all $\theta$ under consideration.

**3. Conditions of regularity.** In this section we shall formulate the restrictions which we impose on $f$, the estimates, and the sequential process. They are intended to be such as will be satisfied in most cases of statistical interest. No doubt they can be weakened, but the author has decided against attempting to do so here. The list may seem long for two reasons. Seldom in the literature are the assumptions which, for example, lead to validation of differentiation under the integral sign etc., formulated explicitly. The presence of a sequential procedure means that additional restrictions must be imposed.

In this section we assume that $\theta$ is a single parameter. The case where $\theta$ has more than one component is treated later.

(3.1). *The parameter $\theta$ lies in an open interval $D$ of the real line. $D$ may consist of the entire line or of an entire half-line.*

(3.2). *The derivative $\frac{\partial f}{\partial \theta}$ exists for all $\theta$ in $D$ and almost all $x$. We define $\frac{\partial \log f(x, \theta)}{\partial \theta}$ as zero whenever $f(x, \theta) = 0$; thus $\frac{\partial \log f}{\partial \theta}$ is defined for all $\theta$ in $D$ and almost all $x$. We postulate that $E \frac{\partial \log f(x, \theta)}{\partial \theta} = 0$ and that $E \left( \frac{\partial \log f(x, \theta)}{\partial \theta} \right)^2$ be not zero for all $\theta$ in $D$.*

(3.3). $$E\left(\sum_{i=1}^{n}\left|\frac{\partial \log f\,(x_i\,,\,\theta)}{\partial\theta}\right|\right)^2$$

*exists for all $\theta$ in $D$.*

(3.4). *Let $R_j$, $(j = 1, 2, \cdots)$, be the set of points $(x_1\,, \cdots, x_j)$ in the $j$-dimensional Euclidean space such that*

$$\varphi_i(x_1\,, \cdots, x_i) = 0 \qquad i = 1, 2, \cdots, j-1$$
$$\varphi_j(x_1\,, \cdots, x_j) = 1.$$

*For any integral $j$ there exists a non-negative L-measurable function $T_j(x_1\,, \cdots, x_j)$ such that*

a). $$\left|\theta^*(x_1, \cdots, x_j)\frac{\partial}{\partial\theta}\prod_{i=1}^{j}f(x_i, \theta)\right| < T_j(x_1, \cdots, x_j)$$

*for all $\theta$ in $D$ and almost all $(x_1\,, \cdots, x_j)$ in $R_j$*

b). $$\int_{R_j} T_j(x_1\,, \cdots, x_j)\,dx_1 \cdots dx_j$$

*is finite.*

(3.5). *Let*

$$t_j(\theta) = \int_{R_j}\theta^*(x_1\,, \cdots, x_j)\prod_{i=1}^{j}f(x_i\,, \theta)\,dx_i\,, \qquad (j = 1, 2, \cdots).$$

*We postulate the uniform convergence of the series*

$$\sum_j \frac{dt_j(\theta)}{d\theta}$$

(*the existence of* $\dfrac{dt_j(\theta)}{d\theta}$ *is a consequence of Assumption* (3.4)) *for all $\theta$ in $D$.*

**4. The case of one parameter.** In this section we assume that $f(x, \theta)$ depends on a single parameter $\theta$. In sections 5 and 6 we shall discuss the case when $\theta$ is a vector with more than one component.

We have $E\,\dfrac{\partial \log f(x, \theta)}{\partial\theta} = 0$

by (3.2). Define the chance variable

$$Y_n = \sum_{i=1}^{n}\frac{\partial \log f(x_i\,, \theta)}{\partial\theta}\,.$$

By an argument almost identical with that of [1], Theorem 1, or of Theorem 7.1 below, we have

$$EY_n = 0. \tag{4.1}$$

From Theorem 7.2 below we obtain

$$\sigma^2(Y_n) = EnE\left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2. \tag{4.2}$$

Let $\theta^*$ $(x_1, \cdots, x_n)$ be an estimate of $\theta$ such that

$$E\theta^* = \theta + b(\theta).$$

Then

$$\sum_{j=1}^{\infty} \int_{R_j} \theta^*(x_1, \cdots, x_j) \prod_{i=1}^{j} f(x_i, \theta)\, dx_i = \theta + b(\theta). \tag{4.3}$$

Differentiation of both members of (4.3) with respect to $\theta$ (Assumptions (3.4) and (3.5)) gives

$$E\theta^* Y_n = 1 + \frac{db}{d\theta}. \tag{4.4}$$

From (4.1) it follows that (4.4) gives the covariance between $\theta^*$ and $Y_n$. Hence from (4.2)

$$\sigma^2(\theta^*) \geq \left(1 + \frac{db}{d\theta}\right)^2 \left[EnE\left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2\right]^{-1}. \tag{4.5}$$

When the bias $b(\theta)$ is constant, for example when $b(\theta) \equiv 0$ in case $\theta^*$ is an unbiased estimate, we have from (4.5)

$$\sigma^2(\theta^*) \geq \left[EnE\left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2\right]^{-1}. \tag{4.6}$$

The equality sign in (4.6) will hold if $\theta^*$ may be written as $Z'(\theta)Y_n + Z''(\theta)$, where $Z'$ and $Z''$ are functions of $\theta$. However, $\theta^*$ itself should not be a function of $\theta$ if our argument is to remain valid. The subject is connected with the question of the existence of a sufficient estimate.

Let $f(x, \theta)$ be defined as follows:

$$f(x, \theta) = \theta^x(1 - \theta)^{1-x}, \qquad (x = 0 \text{ or } 1; 0 < \theta < 1).$$

Then

$$\frac{\partial \log f(x, \theta)}{\partial \theta} = \frac{x}{\theta} - \frac{(1 - x)}{(1 - \theta)}, \qquad E\left(\frac{\partial \log f}{\partial \theta}\right)^2 = \frac{1}{\theta(1 - \theta)}.$$

Suppose $\theta^*$ is unbiased. Then $\sigma^2(\theta^*) \geq \theta(1 - \theta)(En)^{-1}$, a result first given by Girshick [3] under unspecified regularity conditions.

Let the functions $\varphi_1, \varphi_2, \cdots$ be such that $n(\omega)$ is a constant. We are then dealing with samples of fixed size. The result (4.5) is then given in [2], p. 480, under different conditions of regularity.

**5. Regularity conditions for the case when $\theta$ has more than one component.** We suppose that $\theta = (\theta_1, \cdots, \theta_l)$ and that simultaneous estimates

$\theta_1^*(x_1, \cdots, x_n), \cdots, \theta_l^*(x_1, \cdots, x_n)$ of the components of $\theta$ are under discussion. In the sequel we shall limit ourselves to the case when these estimates are all unbiased.

We postulate the following regularity conditions which are sufficient to validate section 6:

(5.1). *The covariance matrix of the estimates $\theta_1^*, \cdots, \theta_l^*$ is non-singular for all $\theta$ in D (this time D is an open interval of the l-dimensional parameter space).*

(5.2). *The conditions of section 3 are satisfied for each $\theta_i$ and $\theta_i^*$ $(i = 1, \cdots, l)$.*

**6. The ellipsoid of concentration when $\theta$ has more than one component.** Let

$$\theta = (\theta_1, \cdots, \theta_l).$$

We shall first describe briefly the result of Cramér [4] which refers to samples of fixed size $n > l$. Let $\theta_i^*(x_1, \cdots, x_n)$ be an unbiased estimate of $\theta_i$, $(i = 1, \cdots, l)$. Let $\| \lambda_{ij} \|$ be the non-singular covariance matrix of the $\theta_i^*$, and let $\| \lambda^{ij} \|$ be its inverse. The "ellipsoid of concentration" in the space of points $(k_1, \cdots, k_l)$ is defined as

$$\sum_{i,j=1}^{l} \lambda^{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2. \tag{6.1}$$

If a unit mass be distributed uniformly over this ellipsoid it will have the point $(\theta_1, \cdots, \theta_l)$ as its center of gravity and $\lambda_{ij}$ as its product of inertia about the corresponding axes. Cramér proves that, subject to certain regularity conditions, there is a fixed ellipsoid

$$\sum_{i,j=1}^{l} \mu_{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2 \tag{6.2}$$

where

$$\mu_{ij} = nE\left(\frac{\partial \log f}{\partial \theta_i} \frac{\partial \log f}{\partial \theta_j}\right)$$

which is always contained entirely within the concentration ellipsoid of any set of unbiased estimates. The two ellipsoids coincide only under certain conditions, among which is that the $\theta_i^*$ be jointly sufficient estimates of the $\theta_i$.

Let us now consider the sequential procedure of this paper and postulate the regularity conditions of section 5. Let

$$K = \| k_{ij} \|$$

be a matrix with real elements such that $| K | = 1$ and let

$$K^{-1} = \| k^{ij} \|$$

be its inverse. Let

$$\| \theta \| = \left\| \begin{matrix} \theta_1 \\ \cdot \\ \cdot \\ \cdot \\ \theta_l \end{matrix} \right\|, \qquad \| \theta^* \| = \left\| \begin{matrix} \theta_1^* \\ \cdot \\ \cdot \\ \cdot \\ \theta_l^* \end{matrix} \right\|, \qquad \| \psi \| = \left\| \begin{matrix} \psi_1 \\ \cdot \\ \cdot \\ \cdot \\ \psi_l \end{matrix} \right\|$$

be column matrices. Suppose

$$\| \psi \| = K \| \theta \|. \tag{6.3}$$

Then

$$\| \theta \| = K^{-1} \| \psi \|. \tag{6.4}$$

Define

$$\| \psi^* \| = \begin{Vmatrix} \psi_1^* \\ \cdot \\ \cdot \\ \cdot \\ \psi_l^* \end{Vmatrix} = K \| \theta^* \|.$$

From section 4 we have

$$EnE\left(\frac{\partial \log f(x, \theta)}{\partial \psi_1}\right)^2 \geq [\sigma^2(\psi_1^*)]^{-1} \tag{6.5}$$

where the differentiation by which $\frac{\partial \log f}{\partial \psi_1}$ is obtained is performed with $\psi_2, \cdots, \psi_l$ held constant. Consider the last $(l - 1)$ rows of $K$ as fixed and $(k_{11}, k_{12}, \cdots, k_{1l})$ as free to vary subject only to the restriction that $| K | = 1$. The left member of (6.5) is then a fixed quantity, while the right member is a function of the first row of $K$. The inequality (6.5) must remain valid for all admissible $(k_{11}, \cdots, k_{1l})$. Hence (6.5) will remain valid if the right member of (6.5) is replaced by its maximum with respect to $(k_{11}, \cdots, k_{1l})$. We shall obtain this maximum and find that (6.5) then implies a result about the minimal ellipsoid of concentration.

The problem is therefore to minimize $\sigma^2(\psi_1^*)$. Now

$$\sigma^2(\psi_1^*) = \sum_{i,j} \lambda_{ij} k_{1i} k_{1j}. \tag{6.6}$$

The family of ellipsoids in the space of $(k_{11}, \cdots, k_{1l})$

$$\sum_{i,j} \lambda_{ij} k_{1i} k_{1j} = c, \tag{6.7}$$

where $c$ is a running parameter, has all centers located at the origin. Let

$$(k_{11}^0, \cdots, k_{1l}^0)$$

be the sought-for maximizing values of $(k_{11}, \cdots, k_{1l})$. From the definitions of $K$ and $K^{-1}$ we have

$$\sum_i k^{i1} k_{1i} = 1 \tag{6.8}$$

where $(k^{11}, k^{21}, \cdots, k^{l1})$ are constants. It follows that the minimum value $c_0$ of $\sigma^2(\psi_1^*)$ is such that the ellipsoid

$$\sum_{i,j} \lambda_{ij} k_{1i} k_{1j} = c_0 \tag{6.9}$$

is tangent to the hyperplane (6.8) at the point $(k_{11}^0, \cdots, k_{1l}^0)$. Now the tangent plane to (6.9) at this point is given by

$$\sum_{i,j} \lambda_{ij} k_{1i}^0 k_{1j} = c_0. \tag{6.10}$$

From (6.8) and (6.10) we obtain

$$c_0 k^{j1} = \sum_i k^0_{1i} \lambda_{ij}, \qquad (j = 1, \cdots, l). \tag{6.11}$$

Hence

$$c_0 \sum_i \lambda^{ij} k^{i1} = k^0_{1j}, \qquad (j = 1, \cdots, l) \tag{6.12}$$

from which

$$c_0 \sum_{i,j} \lambda^{ij} k^{i1} k^{j1} = 1. \tag{6.13}$$

We have

$$\begin{gathered} \frac{\partial \log f}{\partial \psi_1} = \sum_i k^{i1} \frac{\partial \log f}{\partial \theta_i} \\ \left(\frac{\partial \log f}{\partial \psi_1}\right)^2 = \sum_{i,j} k^{i1} k^{j1} \frac{\partial \log f}{\partial \theta_i} \frac{\partial \log f}{\partial \theta_j}. \end{gathered} \tag{6.14}$$

From (6.5), (6.13), (6.14), and the definition of $c_0$ we conclude that

$$\sum_{i,j} \mu'_{ij} k^{i1} k^{j1} \geq \sum_{i,j} \lambda^{ij} k^{i1} k^{j1} \tag{6.15}$$

where

$$\mu'_{ij} = EnE\left(\frac{\partial \log f}{\partial \theta_i} \frac{\partial \log f}{\partial \theta_j}\right). \tag{6.16}$$

We may restate (6.15) as follows: The concentration ellipsoid

$$\sum_{i,j} \lambda^{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2 \tag{6.17}$$

of the unbiased estimates $\theta_1^*, \cdots, \theta_l^*$ always contains within itself the ellipsoid

$$\sum_{i,j} \mu'_{ij}(k_i - \theta_i)(k_j - \theta_j) = l + 2 \tag{6.18}$$

where the $\mu'_{ij}$ are defined by (6.16).

The question of the coincidence of the two ellipsoids is connected with the question of the existence of sufficient estimates. It may be difficult to state any general results about the concentration ellipsoid of biased estimates without postulating some relationships among the biases and/or their derivatives.

**7. On Wald's equation and related results in sequential analysis.** In section 4 we referred to a proof by Blackwell [1] of an equation due to Wald [5] which is fundamental in the Wald theory of sequential tests of statistical hypotheses. Here we shall give a perhaps simpler proof of this equation, and then prove several new and related results of general interest for sequential analysis.

The results of Theorems 7.2 and 7.3 below can be obtained by differentiation of Wald's fundamental identity of sequential analysis ([6], [7]). However, the conditions under which we obtain these results are less stringent than any so far found sufficient to establish the identity and the validity of differentiating it. Theorem 7.4 and its corollaries refer to sequential processes where the chance variables may have different distributions or even be dependent. In the future we hope to return to the question of finding all central moments of $Z_n$, the problem of generalizing the fundamental identity, and related questions.

For Theorems 7.1, 7.2, and 7.3 we shall assume a chance variable $X$ whose cumulative distribution function $F(x)$ is subject only to whatever restrictions may be explicitly imposed on it in each theorem. We assume the existence of a general sequential process such as is described above, which is subject only to such restrictions as may be explicitly formulated in each theorem. The sequential process of course defines the chance variable $n$. Let $x_1, x_2, \cdots$ be successive independent observations on $X$. We define $Z_n = \sum_{i=1}^{n} x_i$. If $E(X)$ and $\sigma^2(X)$ exist we shall denote them by $w$ and $\sigma^2$, respectively.

THEOREM 7.1 (Wald [5], Blackwell [1]). *Suppose $w$ and $En$ exist. Then*

$$E(Z_n - nw) = 0. \tag{7.1}$$

The following theorem, which is a sort of partial converse of Theorem 7.1, is proved concomitantly with Theorem 7.1:

THEOREM 7.1.1. *If $EZ_n$ exists, and if either $P\{X > 0\} = 0$ or $P\{X < 0\} = 0$, then $w$ and $En$ both exist, and*

$$EZ_n = wEn.$$

Actually the same proof suffices for a somewhat stronger form of Theorem 7.1.1:

THEOREM 7.1.2. *If $EZ_n$ exists, and if*

$$E(X_i \mid n = j) \geq 0 \qquad \text{(or } \leq 0\text{)}$$

*for all positive integral $j$ such that $P\ \{n = j\} \neq 0$, and all $i \leq j$, then $w$ and $En$ both exist, and*

$$EZ_n = wEn\ .$$

THEOREM 7.2. *If $E\left(\sum_{i=1}^{n} \mid x_i - w \mid\right)^2$ exists, then $\sigma^2$ and $En$ both exist, and*

$$E(Z_n - nw)^2 = \sigma^2 En\ . \tag{7.2}$$

We have

$$\begin{aligned} E(Z_n - nw) = E\left(\sum_{i=1}^{n} (x_i - w)\right) &= \sum_{j=1}^{\infty} \int_{R_j} \left(\sum_{i=1}^{j} (x_i - w)\right) \prod_{i=1}^{j} dF(x_i) \\ &= \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} (x_j - w) \prod_{m=1}^{m=i} dF(x_m). \end{aligned} \tag{7.3}$$

Also

$$\sum_{i=j}^{\infty}\int_{R_i}(x_j - w)\prod_{m=1}^{m=i} dF(x_m) = P\{n \geq j\}E(x_j - w) = 0. \tag{7.4}$$

Hence

$$\sum_{j=1}^{\infty}\sum_{i=j}^{\infty}\int_{R_i}(x_j - w)\prod_{m=1}^{m=i} dF(x_m) = 0. \tag{7.5}$$

From this (7.1) follows.

Suppose now that the conditions of Theorem 7.2 are fulfilled. We have

$$\begin{aligned} E(Z_n - nw)^2 &= \sum_{j=1}^{\infty}\int_{R_j}\left(\sum_{i=1}^{j}(x_i - w)\right)^2\prod_{m=1}^{m=j} dF(x_m) \\ &= \sum_{j=1}^{\infty}\sum_{i=j}^{\infty}\int_{R_i}(x_j - w)^2\prod_{m=1}^{m=i} dF(x_m) \\ &\quad + 2\sum_{j=2}^{\infty}\sum_{s=1}^{j-1}\sum_{i=j}^{\infty}\int_{R_i}(x_s - w)(x_j - w)\prod_{m=1}^{m=i} dF(x_m). \end{aligned} \tag{7.6}$$

Let $s < j$ be any two positive integers. Then

$$\sum_{i=j}^{\infty}\int_{R_i}(x_s - w)(x_j - w)\prod_{m=1}^{m=i} dF(x_m) = 0. \tag{7.7}$$

Hence

$$\sum_{j=2}^{\infty}\sum_{s=1}^{j-1}\sum_{i=j}^{\infty}\int_{R_i}(x_s - w)(x_j - w)\prod_{m=1}^{m=i} dF(x_m) = 0. \tag{7.8}$$

In a similar manner we obtain

$$\sum_{i=j}^{\infty}\int_{R_i}(x_j - w)^2\prod_{m=1}^{m=i} dF(x_m) = \sigma^2 P\{n \geq j\}. \tag{7.9}$$

From (7.6), (7.8), and (7.9) it therefore follows that

$$E(Z_n - nw)^2 = \sigma^2\sum_{j=1}^{\infty}P\{n \geq j\} = \sigma^2\sum_{j=1}^{\infty} jP\{n = j\} = \sigma^2 En \tag{7.10}$$

which is the desired result.

It remains to prove the validity of rearranging the series in (7.3) and (7.6). First, we have

$$\sum_{i=j}^{\infty}\int_{R_i}|\, x_j - w \,|\prod_{m=1}^{m=i} dF(x_m) = P\{n \geq j\}E\,|\, X - w \,|. \tag{7.11}$$

Hence it follows that

$$(7.12)\qquad \begin{aligned} \sum_{j=1}^{\infty}\sum_{i=j}^{\infty}\int_{R_i} | x_j - w | \prod_{m=1}^{m=i} dF(x_m) &= \sum_{j=1}^{\infty} P\{ n \geq j\} E \mid X - w \mid \\ &= E \mid X - w \mid \sum_{j=1}^{\infty} jP\{n = j\} = E \mid X - w \mid En. \end{aligned}$$

This justifies the rearrangement of terms in the series in (7.3). Second, the series (7.6) is dominated by the series

$$(7.13)\qquad \begin{aligned} &\sum_{j=1}^{\infty}\sum_{i=j}^{\infty}\int_{R_i} (x_j - w)^2 \prod_{m=1}^{m=i} dF(x_m) \\ &\qquad + 2\sum_{j=2}^{\infty}\sum_{s=1}^{j-1}\sum_{i=j}^{\infty}\int_{R_i} | x_s - w |\cdot| x_j - w | \prod_{m=1}^{m=i} dF(x_m) \end{aligned}$$

all of whose terms are positive. The series (7.13) converges because

$$(7.14)\qquad E\left(\sum_{i=1}^{n} | x_i - w |\right)^2 < +\infty.$$

Hence the rearrangement of the series (7.6) is valid.

In the sequel we require certain sets $\mathrm{R}'_j(j = 1, 2, \cdots)$ which we shall define now. Let $R^*_{ij}$, $i \leq j$, be the totality of all points $(x_1, \cdots, x_j)$ such that

$$(7.15)\qquad (x_1, \cdots, x_i) \in R_i .$$

Let $R^j$ be the $j$-dimensional Euclidean space. Then

$$(7.16)\qquad R'_j = R^j - \sum_{i=1}^{j} R^*_{ij} .$$

We shall now prove:

THEOREM 7.3. *Suppose that* $E\left[\sum_{i=1}^{n} | x_i - w |\right]^3$ *and* $En\left[\sum_{i=1}^{n} | x_i - w |\right]$ *exist.*[2] *Then*

$$(7.17)\qquad E(Z_n - nw)^3 = w_3 En + 3\sigma^2 En(Z_n - nw)$$

*where*

$$w_3 = E(X - w)^3$$

*exists.*

[2] The author has succeeded in proving that the existence of $E\left[\sum_{i=1}^{n} | x_i - w |\right]^3$ implies the existence of $E\left[n \sum_{i=1}^{n} | x_i - w |\right]$. The proof will be published subsequently in connection with other results.

**Proof:** We have

$$E(Z_n - nw)^3 = \sum_{j=1}^{\infty} \int_{R_j} \left[ \sum_{i=1}^{j} (x_i - w) \right]^3 \prod_{m=1}^{j} dF(x_m)$$

$$\begin{aligned} &= \sum_{j=1}^{\infty} \int_{R_j} \sum_{i=1}^{j} (x_i - w)^3 \prod_{m=1}^{j} dF(x_m) \\ &+ 3 \sum_{j=2}^{\infty} \int_{R_j} \sum_{i=2}^{j} \sum_{s=1}^{i-1} (x_s - w)(x_i - w)^2 \prod_{m=1}^{j} dF(x_m) \\ &+ 3 \sum_{j=2}^{\infty} \int_{R_j} \sum_{i=2}^{j} \sum_{s=1}^{i-1} (x_s - w)^2 (x_i - w) \prod_{m=1}^{j} dF(x_m) \\ &+ 6 \sum_{j=3}^{\infty} \int_{R_j} \sum_{i=3}^{j} \sum_{s=2}^{i-1} \sum_{t=1}^{s-1} (x_t - w)(x_s - w)(x_i - w) \prod_{m=1}^{j} dF(x_m). \end{aligned} \tag{7.18}$$

Considering the first term in the right member of (7.18), it follows that

$$\begin{aligned} &\sum_{j=1}^{\infty} \int_{R_j} \left[ \sum_{i=1}^{j} (x_i - w)^3 \right] \prod_{m=1}^{j} dF(x_m) \\ &= \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \int_{R_j} (x_i - w)^3 \prod_{m=1}^{j} dF(x_m) \\ &= \sum_{i=1}^{\infty} w_3 P\{n \geq i\} \\ &= \sum_{i=1}^{\infty} i w_3 P\{n = i\} = w_3 En. \end{aligned} \tag{7.19}$$

All the rearrangements of terms in the operations involved in the proof of Theorem 7.3 are legitimate because the various series are absolutely convergent.

As for the second term in the right member of (7.18), we have

$$\begin{aligned} &\sum_{j=2}^{\infty} \int_{R_j} \sum_{i=2}^{j} \sum_{s=1}^{i-1} (x_s - w)(x_i - w)^2 \prod_{m=1}^{j} dF(x_m) \\ &= \sum_{s=1}^{\infty} \sum_{i=s+1}^{\infty} \sum_{j=i}^{\infty} \int_{R_j} (x_s - w)(x_i - w)^2 \prod_{m=1}^{j} dF(x_m) \\ &= \sigma^2 \sum_{s=1}^{\infty} \sum_{i=s+1}^{\infty} \int_{R'_{i-1}} (x_s - w) \prod_{m=1}^{i-1} dF(x_m) \\ &= \sigma^2 \sum_{s=1}^{\infty} \sum_{i=s}^{\infty} \int_{R'_i} (x_s - w) \prod_{m=1}^{i} dF(x_m). \end{aligned} \tag{7.20}$$

We now operate on $En(Z_n - nw)$, and obtain

$$\begin{aligned} En(Z_n - nw) &= \sum_{j=1}^{\infty} \int_{R_j} j \sum_{i=1}^{j} (x_i - w) \prod_{m=1}^{j} dF(x_m) \\ &= \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} i(x_j - w) \prod_{m=1}^{i} dF(x_m). \end{aligned} \tag{7.21}$$

We observe that

$$\begin{aligned}
&\sum_{i=j}^{\infty}\int_{R_i} i(x_j - w)\prod_{m=1}^{i} dF(x_m) \\
(7.22)\qquad &= j\sum_{i=j}^{\infty}\int_{R_i}(x_j - w)\prod_{m=1}^{i} dF(x_m) \\
&+ \sum_{s=j+1}^{\infty}\sum_{i=s}^{\infty}\int_{R_i}(x_j - w)\prod_{m=1}^{i} dF(x_m).
\end{aligned}$$

To evaluate the left member of (7.22), we proceed as follows: It is easy to see that

$$(7.23)\qquad \sum_{i=j}^{\infty}\int_{R_i}(x_j - w)\prod_{=1}^{i} dF(x_m) = 0.$$

Moreover, when $s > j$,

$$(7.24)\qquad \sum_{i=s}^{\infty}\int_{R_i}(x_j - w)\prod_{m=1}^{i} dF(x_m) = \int_{R'_{s-1}}(x_j - w)\prod_{m=1}^{s-1} dF(x_m).$$

Hence

$$(7.25)\qquad \sum_{i=j}^{\infty}\int_{R_i} i\,(x_j - w)\prod_{m=1}^{i} dF(x_m) = \sum_{s=j}^{\infty}\int_{R'_s}(x_j - w)\prod_{m=1}^{s} dF(x_m).$$

Therefore

$$(7.26)\qquad En(Z_n - nw) = \sum_{j=1}^{\infty}\sum_{s=j}^{\infty}\int_{R'_s}(x_j - w)\prod_{m=1}^{s} dF(x_m).$$

It remains now to consider the third term of the right member of (7.18). We have

$$\begin{aligned}
&\sum_{j=2}^{\infty}\int_{R_j}\sum_{i=2}^{j}\sum_{s=1}^{i-1}(x_s - w)^2(x_i - w)\prod_{m=1}^{j} dF(x_m). \\
(7.27)\qquad &= \sum_{s=1}^{\infty}\sum_{i=s+1}^{\infty}\sum_{j=i}^{\infty}\int_{R_j}(x_s - w)^2(x_i - w)\prod_{m=1}^{j} dF(x_m).
\end{aligned}$$

Now, suppose that in the expression

$$(7.28)\qquad V_{sij} = \int_{R_j}(x_s - w)^2(x_i - w)\prod_{m=1}^{j} dF(x_m)$$

where $j \geq i > s$, we integrate with respect to all $x_m$ for which $m \geq i$. Then it is not difficult to see that

$$(7.29)\qquad \sum_{j=i}^{\infty} V_{sij} = 0$$

for all $s$ and $i$ such that $1 \leq s < i$. Hence from (7.27)

$$(7.30)\qquad \sum_{j=2}^{\infty}\int_{R_j}\sum_{i=2}^{j}\sum_{s=1}^{i-1}(x_s - w)^2(x_i - w)\prod_{m=1}^{j} dF(x_m) = 0.$$

In a similar way it is shown that the fourth term of the right member of (7.18) is zero.

The desired result (7.17) is a direct consequence of (7.18), (7.19), (7.20), (7.26), and (7.30).

Consider now an infinite sequence of chance variables $x_1, x_2, \cdots$, which need not have the same distribution and which may be dependent (in which case they must satisfy the obvious consistency relationships). We take successive observations on these chance variables and define a sequential process as above, which is subject only to such restrictions as we shall explicitly state. Let $Z_n$ maintain its previous definition.

THEOREM 7.4. *Suppose that*

$$\nu_i = E(X_i \mid n \geq i) \tag{7.31}$$

*exists for all positive integral i for which* $P\{n \geq i\} \neq 0$. *In those cases write*

$$\nu_i' = E(|X_i - \nu_i| \,\big|\, n \geq i). \tag{7.32}$$

*Suppose also that the series*

$$\sum_{i=1}^{\infty} (\nu_1' + \cdots + \nu_i')P\{n = i\} \tag{7.33}$$

*converges. Then*

$$E\left[Z_n - \sum_{i=1}^{n} \nu_i\right] = 0. \tag{7.34}$$

It is regrettable but unavoidable that the mean values $\nu_i$ and $\nu_1'$ entering into (7.33) and (7.34) be conditional. The fundamental reason is that the sequential process may drastically modify the distribution of dependent chance variables, so that their distribution for our purposes can only be considered in conjunction with the sequential process itself. Consider the following example:

$$P\{X_1 = -1\} = \tfrac{1}{2}, \qquad P\{X_1 = 1\} = \tfrac{1}{2}$$

$$P\{X_2 = -2 \mid X_1 = -1\} = \tfrac{1}{2}$$

$$P\{X_2 = -1 \mid X_1 = -1\} = \tfrac{1}{2}$$

$$P\{X_2 = 1 \mid X_1 = 1\} = \tfrac{1}{2}$$

$$P\{X_2 = 2 \mid X_1 = 1\} = \tfrac{1}{2}.$$

We have $E(X_2) = 0$. Suppose we define the following sequential process: If $X_1 = -1, n = 1$, and if $X_1 = 1, n = 2$. It is then clear that for our purposes $X_2$ can take no negative values and the fact that $E(X_2) = 0$ is of no use to us.

If, however, the chance variables $X_1, X_2, \cdots$ are independent, this difficulty disappears, and we have the following.

COROLLARY 1 TO THEOREM 7.4. *If the chance variables $X_1, X_2, \cdots$ are independent, we have Theorem* 7.4 *with $\nu_i = E(X_i)$, and $\nu_i' = E \mid X_i - \nu_i \mid$.*

If further all the $X_i$ have the same distribution, we see that Theorem 7.1 is a special case of Theorem 7.4, since the convergence of the series (7.33) is then a consequence of the existence of $w$ and $En$. From this argument we see, however, that it is not necessary that all the $X_i$ have the same distribution, and we may write the following generalization of Theorem 7.1:

COROLLARY 2 TO THEOREM 7.4. *Let the $X_i$ be independent with, in general, different distributions. Suppose, however, that all $\nu_i$ are equal, and all $\nu_i'$ are equal, except perhaps for those $i$ such that $P\{n \geq i\} = 0$. Suppose further that $En$ exists. Then* (7.1) *holds.*

Among possible fields of application of Theorem 7.4 are sequential tests of composite statistical hypotheses, and the random walk of a particle governed by probability distributions which are functions of time and the position of the particle. The extension of this theorem to vector chance variables is straightforward. The extension to higher moments may present difficulties. We hope to return to some of these questions in the future.

PROOF OF THEOREM 7.4. This is very elementary. We have

$$\begin{aligned} E\left(Z_n - \sum_{i=1}^{n} \nu_i\right) &= \sum_{j=1}^{\infty} \int_{R_j} \left[\sum_{i=1}^{j} (x_i - \nu_i)\right] dF(x_1, \cdots, x_j) \\ &= \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} (x_j - \nu_j)\, dF(x_1, \cdots, x_i). \\ &= \sum_{j=1}^{\infty} P\{n \geq j\} E(X_j - \nu_j \mid n \geq j) = 0. \end{aligned} \tag{7.35}$$

The rearrangement of the series is valid because

$$\begin{aligned} \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \int_{R_i} \mid x_j - \nu_j \mid dF(x_1, \cdots, x_i) &= \sum_{j=1}^{\infty} \nu_j' P\{n \geq j\} \\ &= \sum_{j=1}^{\infty} (\nu_1' + \cdots + \nu_j') P\{n = j\} \end{aligned} \tag{7.36}$$

which converges by (7.33).

## REFERENCES

[1] DAVID BLACKWELL, "On an equation of Wald," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 84–87.

[2] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946.

[3] M. A. GIRSHICK, FREDERICK MOSTELLER, AND L. J. SAVAGE, "Unbiased estimates for certain binomial sampling problems," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 13–23.

[4] H. Cramér, "A contribution to the theory of statistical estimation," *Skandinavisk Aktuarietidskrift*, Vol. 29 (1946), pp. 85-94.

[5] A. Wald, "Sequential tests of statistical hypotheses," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 117-186.

[6] A. Wald, "On cumulative sums of random variables," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 283-296.

[7] A. Wald, "Differentiation under the expectation sign of the fundamental identity in sequential analysis," *Annals of Math. Stat.*, Vol. 17 (1946).

[8] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, Vol. 37, No. 3 (Sept., 1945), pp. 81-91.

[9] A. C. Aitken and H. Silverstone "On the estimation of statistical parameters," *Proc. Roy. Soc. Edinburgh*, Vol. 61 (1941), pp. 56-62.

# ESTIMATION OF LINEAR FUNCTIONS OF CELL PROPORTIONS

By John H. Smith

*Bureau of Labor Statistics*

**Summary.** In this article certain contributions are made to the theory of estimating linear functions of cell proportions in connection with the methods of (1) least squares, (2) minimum chi-square, and (3) maximum likelihood. Distinctions among these three methods made by previous writers arise out of (1) confusion concerning theoretical vs. practical weights, (2) neglect of effects of correlation between sampling errors, and (3) disagreement concerning methods of minimization. Throughout the paper the equivalence of these three methods from a practical point of view has been emphasized in order to facilitate the integration and adaptation of existing statistical techniques. To this end:

1. The method of least squares as derived by Gauss in 1821–23 [6, pp. 224–228] in which weights in theory are chosen so as to minimize sampling variances is herein called the ideal method of least squares and the theoretical estimates are called ideal linear estimates. This approach avoids confusion between practical approximations and theoretical exact weights.

2. The ideal method of least squares is applied to uncorrelated linear functions of correlated sample frequencies to determine the appropriate quantity to minimize in order to derive ideal linear estimates in sample-frequency problems. This approach leads to a sum of squares of standardized uncorrelated linear functions of sampling errors in which statistics are to be substituted in numerators.

3. A new elementary method is used to reduce the sum of squares in (2)—before substitution of statistics—to Pearson's expression for chi-square. In this result, obtained without approximation, appropriate substitution of statistics shows that the denominators of chi-square should be treated as constant parameters in the differentiation process in order to minimize chi-square in conformity with the ideal method of least squares.

4. The ideal method of minimum chi-square, derived in (3) as the sample-frequency form of the ideal method of least squares, yields ideal linear estimates in terms of the unknown parameters in the denominators of chi-square. When these parameters are estimated by successive approximations in such a way as to be consistent with statistics based on them, it is shown that the method of minimum chi-square leads to maximum likelihood statistics.

5. An iterative method which converges to maximum likelihood estimates is developed for the case in which observations are cross-classified and first order totals are known. In comparison with Deming's asymptotically efficient statistics, it is shown that, in a certain sense, maximum likelihood statistics are suuerior for any given value of $n$—especially in small samples.

6. The method of proportional distribution of marginal adjustments is de-

veloped. This method yields estimates of expected cell frequencies whose efficiency is 100 per cent when universe cell frequencies are proportional—a condition closely approximated in most practical surveys for which first order totals are available from complete censuses. Whether this favorable condition is satisfied or not, the method yields results which are easy to interpret and it has many computational advantages from the point of view of economy of time and effort.

Throughout the article discussion is confined to the estimation of parameters whose relationships to cell proportions are linear. However, most of the results can be extended to the case of non-linear relationships, the necessary qualifications being similar to those in curve-fitting problems when the function to be fitted is not linear in its parameters. In this case, of course, least squares estimates are not linear estimates. In particular, obvious extensions of the general proofs in sections 5 and 6 make them applicable to the non-linear case. Thus even when relationships are non-linear, it can be shown that the method of minimum chi-square is the sample-frequency form of the method of least squares which leads (by means of appropriate successive approximations) to maximum likelihood statistics in sample-frequency problems. This principle which establishes the equivalence of the methods of least squares, minimum chi-square, and maximum likelihood greatly facilitates the integration and adaptation of existing techniques developed in connection with these important methods of estimation.

**1. Introduction.** This article deals with problems of statistical estimation in which the parameters to be estimated are cell proportions or linear functions of them. A simple illustration of this type of problem is that of estimating $p$, the proportion of white men in a population classified by race and sex. Fom a sample of $n$ persons selected at random from such a population, the desired proportion can be estimated by simply taking the sample proportion of white men as an estimate of the corresponding cell proportion in the population or universe. This estimate is unbiased for all possible values of $p$ and its sampling variance is $p(1 - p)/n$—assuming, for simplicity, that sampling is done with replacements. Whether a more accurate unbiased estimate of $p$ can be derived depends on whether or not any other relevant information concerning the cell proportions in the universe is available. For example, it may be known that all of the white portion of the population is composed of married couples so that in the universe the number of white men is exactly equal to the number of white women. This knowledge implies that half the proportion of whites provides an unbiased estimate of $p$ which is far more accurate than the sample proportion of white men. In fact, the sampling variance of half the proportion of whites is equal to $(2p)(1 - 2p)/4n$—less than half the sampling variance of the proportion of white men.

The term *ideal linear estimate* will be used to refer to any statistic which satisfies the criteria of estimation implied by the foregoing discussion—that is, an ideal linear esimate is any estimate which (1) is a linear function of the sample observations; (2) is recognizable as unbiased by the research worker; and (3) has minimum sampling variance among estimates which have properties (1) and (2). These important criteria of estimation will now be stated in more technical language.

Let $n_1$, $n_2$, and $n_3$ represent the number of (1) white men, (2) white women and (3) non-white persons, respectively, in samples of $n$ persons. Since any linear function with a constant term can be reduced to the homogeneous form by adding an appropriate multiple of the identity

$$n_1 + n_2 + n_3 - n \equiv 0, \tag{1.1}$$

it is possible, without loss of generality, to confine attention to linear estimates of the form

$$T = a_1n_1 + a_2n_2 + a_3n_3, \tag{1.2}$$

which are recognizable as unbiased. In this example, the research worker is assumed to know that the cell proportions in the universe are

$$p_1, p_2, p_3 = p, p, 1 - 2p. \tag{1.3}$$

Hence, absence of bias implies that the expected value of $T$

$$\begin{aligned} E(T) &= a_1np_1 + a_2np_2 + a_3np_3 \\ &= (a_1 + a_2 - 2a_3)np + na_3 \end{aligned} \tag{1.4}$$

is identically equal to $p$; in other words, that

$$n(a_1 + a_2 - 2a_3) - 1 = 0, \tag{1.5}$$

and

$$na_3 = 0.$$

The ideal linear esimate is derived by finding values of $a_1$, $a_2$, and $a_3$ which minimize the sampling variance of $T$ subject to equations (1.5) as side conditions.[1] In this way it can be shown that half the sample proportion of whites is actually the ideal linear estimate of $p$. For more general problems, the process of minimization of sampling variances with the aid of Lagrange multipliers involves expressions which are complicated algebraically. For this reason it is usually easier to derive ideal linear estimates of parameters which are linear functions of cell proportions by the ideal method of least squares which is presented in section 4.

Like other least squares estimates, an ideal linear estimate of a linear function of cell proportions depends on ideal least squares weights. Since these weights

[1] In this example, it is possible to solve equations (1.5) for $a_2$ in terms of $a_1$, drop subscripts, and substitute in the formula for the sampling variance of $T$ to obtain a quadratic in $a$ to be minimized.

are, in general, functions of variances and covariances of sample frequencies, the theoretical connotation of the term "ideal" makes it preferable to other terms such as "optimum" and "best." In this connection it should be emphasized that (1) the sampling variance of linear estimates is insensitive to small errors in estimating ideal weights, and (2) the process of deriving practical approximations to ideal linear estimates automatically provides maximum likelihood estimates of the ideal weights. Thus the estimation of weights is perfectly objective and the best practical approximations to ideal linear estimates are expressed in terms of sample observations. This degree of objectivity is rare in statistical estimation as a brief consideration of regression problems will illustrate.

In ordinary regression problems, the ideal weights are inversely proportional to error variances. It is usually necessary to draw upon past experience to estimate relative weights because satisfactory estimates of error variances are rarely available in terms of sample observations. From the present point of view, the widespread use of equal weights implies the *subjective* "assumption" that all error variances are equal. (Maximum likelihood estimates of regression coefficients require, in addition, the even more subjective assumption of normality.) In spite of these (usually implicit) subjective assumptions, discussions of optimum properties of least squares regression coefficients based on *ideal* weights in terms of *unknown parameters* are highly commendable because (1) sampling variance is not very sensitive to small errors in weights and (2) properties of theoretical ideal linear estimates furnish a simple basis for discussion of the properties of practical statistics based on any reasonably good approximations to the exact ideal weights. In any case, it is important to know what the ideal weights are in terms of unknown parameters because research workers can make better estimates if they know what quantities should be estimated than they could otherwise.

**2. Estimation of a single parameter.** In sample-frequency problems, least squares weights are rarely given explicitly or even implied by information available to the research worker. Since the hypothetical example used in Section 1 is a trivial special case from this point of view, a more realistic example is presented in this section. Since the biological interpretation of this problem is presented in detail in all but the first of the many editions of Fisher's well-known book [3] it is sufficient here to consider only the statistical problem. The four cell proportions are

$$p_1, p_2, p_3, p_4 = (2 + \theta)/4, (1 - \theta)/4, (1 - \theta)/4, \theta/4, \tag{2.1}$$

and the parameter $\theta$ is to be estimated from the set of sample frequencies

$$n_1, n_2, n_3, n_4 = 1997, 906, 904, 32, \tag{2.2}$$

obtained in a sample of $n = 3839$ selected at random from an infinite universe. Fisher considers five different statistics—$T_1$, $T_2$, $T_3$, $T_4$, and $T_5$—so it will be convenient to use the symbol $T_6$ for the ideal linear estimate. Consider the class of linear unbiased estimates of the form

$$T = a_1n_1 + a_2n_2 + a_3n_3 + a_4n_4\,, \tag{2.3}$$

where absence of bias implies that

$$2a_1 + a_2 + a_3 = 0 \tag{2.4}$$

and

$$a_1 - a_2 - a_3 + a_4 - 4/n = 0.$$

Minimizing the sampling variance of $T$ in equation (2.3) subject to side conditions based on equations (2.4) yields the ideal linear estimate $T_6$ defined by the equation

$$n(1 + 2\theta)T_6 = 3\theta n_1 - 3\theta n_2 - 3\theta n_3 + (4 - \theta)n_4\,. \tag{2.5}$$

The exact sampling variance of $T_6$,

$$\sigma_6^2 = \frac{2\theta(1 - \theta)(2 + \theta)}{n(1 + 2\theta)}\,, \tag{2.6}$$

is used by Fisher as the asymptotic sampling variance of any efficient estimate of $\theta$. The exact sampling variance of the ideal linear estimate is especially appropriate as the asymptotic sampling variance of the maximum likelihood estimate $T_4$ because $T_4$ is the limit of an iterative process designed to estimate $T_6$ as closely as possible from sample data by using successive approximations to $T_6$ for $\theta$ in equation (2.5). The limit of this process (which is, of course, only an approximation to $T_6$) can be obtained by substituting the symbol $T_4$ for both $T_6$ and $\theta$ in equation (2.5) and solving the resulting quadratic equation which can be reduced to

$$nT_4^2 - (n_1 - 2n_2 - 2n_3 - n_4)T_4 - 2n_4 = 0, \tag{2.7}$$

an equation which is identical, except for notation, with Fisher's equation of maximum likelihood of which $T_4$ is the positive solution.

The foregoing result is a comparatively simple illustration of the general principle that the maximum likelihood estimate of any linear function of cell proportions is the limit of an iterative process designed to approximate the corresponding linear estimate as closely as possible by means of sample frequencies. Since the accuracy of estimates of least squares relative weights increases with size of sample, maximum likelihood statistics have, in an asymptotic sense for large samples, the same optimum properties which are possessed in an exact sense (even for small samples) by the corresponding ideal linear estimates. Thus the results obtained by means of the theory of large samples are supported by the approach to estimation problems by means of ideal linear estimates. In addition, the later approach facilitates the integration of available techniques as explained in later sections.

It is true that the optimum properties of maximum likelihood statistics can be presented in terms of the theory of large samples, but the fact that a given method of estimation yields a statistic whose asymptotic sampling variance is a minimum does not imply that the same technique will yield a minimum variance statistic for any given small value of $n$. For example, it is well known that the median is a maximum likelihood estimate of the midpoint of a double exponential universe. Nevertheless, in samples of three observations from such a universe, another statistic—4/9 of the mean plus 5/9 of the median—has greater relative advantage over the median than the median has over the mean.

Fisher's discussion of the relative efficiencies of his five alternative consistent statistics suggests that it is impossible to formulate objective criteria for making choices among alternative statistics such that each statistic will be used whenever its sampling variance is smallest. Consider the sequence of universes generated by letting $\theta$ vary from zero to unity. In general, each value of $\theta$ would determine which of Fisher's five statistics would have smallest sampling variance for that particular universe for any given value of $n$. In comparison with any other single statistic, the statistic $T_4$ would usually have smaller sampling variance, but there are notable exceptions. For example, in the absence of linkage when $\theta$ is equal to one-fourth, the statistic $T_2$ is the ideal linear estimate and its sampling variance is smaller than that of $T_4$—at least for certain small values of $n$. For this reason, Fisher used $T_2$ in preference to $T_4$ as the basis for testing the significance of linkage. The statistic $T_5$—derived by Fisher's method of minimum chi-square—is also of special interest. Fisher's method of minimum chi-square yields statistics which differ from the corresponding maximum likelihood statistics because Fisher considers the denominators as variables in the process of differentiation instead of considering them as unknown parameters to be estimated by identifying them with the corresponding statistics in the numerators *after* differentiation. Arguments of later sections tend to show that the latter method is more appropriate. In this example, it can be shown that if $T_5$ were substituted for the corresponding parameter in the denominators of chi-square (*and treated as a parameter*) the minimization of chi-square with respect to statistics in its numerators would be exactly equivalent to substituting 0.035785, the numerical value of $T_5$ for $\theta$ in equation (2.5) and solving for $T_6$ to obtain 0.035717, a value which is much closer to 0.035712, the numerical value of the maximum likelihood estimate $T_4$ than to Fisher's $T_5$. In problems of estimation chi-square should be minimized in order to obtain efficient statistics—not to obtain a small criterion for testing goodness of fit—and it should be minimized in a manner consistent with this purpose. Whether or not it is possible to derive an even smaller value for a quantity called chi-square should be considered to be irrelevant in either estimation problems or tests of significance. It is difficult to present these ideas in more technical language because it is possible to construct trivial hypothetical universes for which Fisher's method of minimum chi-square provides statistics which are superior in certain respects to the corresponding maximum likelihood statistics. Nevertheless, it seems clear that the ideal linear estimate usually has smaller sampling variance than the maximum likelihood statistic which, in turn, usually has smaller sampling variance than any other given practical statistic. Evidence presented in later sections tends to show that these advantages are more important in small samples than in cases in which the theory of large samples is applicable.

**3. The "ideal" method of least squares.** When sample observations are uncorrelated in successive samples and parameters to be estimated are linear functions of the expected values of the sample observations, the method of least squares yields ideal linear estimates of the parametes provided that the weight of each observation is inversely proportional to its variance in successive samples. Although the minimum sampling variance property among linear unbiased estimates is seldom stressed, this principle of weighting has been presented in connection with the method of least squares for more than a hundred years. In order to emphasize the theoretical nature of weights which depend on variances which are usually unknown in practice and to distinguish the method based on such weights from the more familiar method of least squares with equal weights, the method which yields ideal linear estimates will be called the *ideal method of least squares.*

Discussion of the general problem of estimating linear functions of cell proportions can be facilitated by making use of results obtained by other writers—notably Gauss (as reported by Whittaker and Robinson [6]) and Pearson [4]. According to Whittaker and Robinson, "the first writer to connect the method [of ideal least squares] with the theory of probability was Gauss" [6, p. 224]. In his Theoria Motus proof of 1809, Gauss derived the "most probable value" [6, p. 223] of a parameter (i.e., the statistic which satisfies the criterion now called maximum likelihood) for the case in which sample observations are statistically independent and normally distributed. In his Theoria Combinationis proof of 1821-23, Gauss "abandoned the 'metaphysical' basis" [6, p. 220] of his earlier work and derived the method herein called the ideal method of least squares (without approximation) from the criteria of (1) minimum variance and (2) absence of bias for the case in which "the mean value of [the covariance of a pair of errors] is zero" [6, p. 224]. Since the covariances of *uncorrelated* linear functions are zero whether they are *statistically independent* or not, it follows from the work of Gauss that the ideal method of least squares applied to uncorrelated linear functions of sample frequencies yields ideal linear estimates. In other words, the ideal method of least squares implies the following six steps:

1. From the set of $k + 1$ sample frequencies construct $k$ linear functions which are uncorrelated in successive samples.
2. From each function subtract its expected value in terms of the unknown parameters to find its sampling error.

3. Write the ratio of each sampling error to its own standard error in the form of a fraction.
4. Sum the squares of these standardized uncorrelated sampling errors to obtain a quantity called chi-square.
5. Substitute statistics[2] for the parameters in the *numerators* of chi-square.
6. Minimize the sum of squares of residuals with respect to each statistic in turn (subject to appropriate side conditions in case linear functions not implied in preceding steps are known).

This series of six steps can be summarized by the single statement that the function to minimize is the sum of squares of standardized uncorrelated residuals. Actually this statement is oversimplified because even though sampling errors are both uncorrelated and standardized, the corresponding residuals are, in general, neither standardized nor uncorrelated.

**4. Pearson's expression for chi-square.** As defined by Pearson [4], chi-square is the sum of squares of a set of $k$ standardized uncorrelated linear functions of sampling errors in a set of $k + 1$ correlated sample frequencies. A set of $k$ standardized uncorrelated linear functions can be constructed in an infinite number of ways, but each set can be obtained from any of the others by means of an orthogonal transformation. Thus the sum of squares is the same no matter what set is originally chosen. As his set of standardized uncorrelated linear functions, Pearson chose those determined by the axes of the correlation ellipse for which he gave the required sum of squares in terms of "minors" or cofactors of the correlation determinant of the first $k$ sample frequencies. Pearson reduced this complicated expression to the now familiar form

$$\chi^2 = \sum_{i=1}^{k+1} (n_i - np_i)^2/np_i, \tag{4.1}$$

where $p_i$ is the proportion in the $i$th cell in the universe and $n_i$ is the frequency in the $i$th cell of a sample of $n$ observations selected at random from an infinite universe (or with replacements from a finite universe).

The widespread misunderstanding of the nature of chi-square seems to be based primarily on the facts that

1. Pearson's rule for degrees of freedom is inadequate (see section 5), and
2. Pearson's expression for chi-square can be derived by approximate methods as well as by exact methods.

Pearson's derivation of the expression for chi-square by exact methods is sufficient to show that its derivation by approximate methods involves a paradox in which different sets of approximations offset each other; however, Pearson's article is relatively inaccessible and, in addition, his algabraic reductions involve

[2] It is convenient to call these variable symbols "statistics"; the quantities whose squares are summed, "residuals"; and the whole expression "chi-square," even though, from a certain point of view, these terms are strictly applicable only after the minimization process. This usage should always be clear from its context.

the minors of a general determinant of the $k$th order. For these reasons, the following exact derivation is presented in terms of elementary algebra.

Since the sum of squares is the same for any set of $k$ standardized uncorrelated linear functions of the sampling errors in $k + 1$ correlated frequencies, a set should be chosen for which the algebraic reductions are as easy as possible. From this point of view a satisfactory set, which can be written in any of three forms, is given by

$$\begin{aligned} y_i &= p_i n_{i+} - p_{i+} n_i \\ &= p_i e_{i+} - p_{i+} e_i \\ &= -p_i e_{i-} - (p_i + p_{i+}) e_i \end{aligned} \tag{4.2}$$

where $e_i = n_i - np_i$ and $i+$ and $i-$ refer to classes formed by combining all classes above the $i$th class and below the $i$th class, respectively.

By means of the known variances and covariances of the sample frequencies in expected value form,

$$E(e_i^2) = np_i(1 - p_i), \tag{4.3}$$

and

$$E(e_i e_j) = -np_i p_j\,, \tag{4.4}$$

it can be shown that the variance of $y_i$ is

$$E(y_i) = np_i p_{i+}(p_i + p_{i+}), \tag{4.5}$$

and, by using the third expression in equation (4.2) for $y_i$ and the second for $y_j$, it can be shown that any pair of $y$'s are uncorrelated because

$$E(y_i y_j) = 0, \qquad (i < j). \tag{4.6}$$

Let $z_i$ represent the variable $y_i$ expressed in standard-deviation units. The square of this standardized uncorrelated linear function of correlated sampling errors can be written

$$z_i^2 = \frac{(p_i e_{i+} - p_{i+} e_i)^2}{np_i p_{i+}(p_i + p_{i+})}\,. \tag{4.7}$$

It remains to show that Pearson's expression for chi-square can be obtained by adding the $k$ values of $z_i^2$ in succession. For this purpose it is convenient to define

$$\chi_r^2 = \sum_{i=1}^{r} \frac{e_i^2}{np_i} + \frac{e_{r+}^2}{np_{r+}}\,, \tag{4.8}$$

obtained by combining all classes above the $r$th class.

When $r = k$, the expression in equation (4.8) is the expression to be derived. It remains to show that $\chi_k^2$ is the sum of squares of $k$ standardized uncorrelated linear functions of sampling errors; i.e.,

$$\chi_k^2 = \sum_{i=1}^{k} z_i^2 . \tag{4.9}$$

For the first cell $e_{1+} = -e_1$ and $p_{1+} = 1 - p_1$. Hence $y_1$ reduces to the negative of the error in the first frequency and

$$\begin{aligned} \chi_1^2 &= e_1^2/np_1(1 - p_1) \\ &= e_1^2/np_1 + e_{1+}^2/np_{1+} \qquad (p_{1+} = 1 - p_1), \end{aligned} \tag{4.10}$$

a special case expressed in the required form. The general case is established by showing that

$$\chi_{r-1}^2 + z_r^2 = \chi_r^2 , \tag{4.11}$$

or, alternatively, that

$$\begin{aligned} z_r^2 &= \chi_r^2 - \chi_{r-1}^2 \\ &= e_r^2/np_r + e_{r+}^2/np_{r+} - (e_r + e_{r+})^2/n(p_r + p_{r+}) \\ &= \frac{(p_{r+} e_r^2 + p_r e_{r+}^2)(p_r + p_{r+}) - p_r p_{r+}(e_r^2 + 2e_r e_{r+} + e_{r+}^2)}{np_r p_{r+}(p_r + p_{r+})} \\ &= \frac{p_r^2 e_{r+}^2 - 2p_r p_{r+} e_r e_{r+} + p_{r+}^2 e_r^2 \equiv (p_r e_{r+} - p_{r+} e_r)^2}{np_r p_{r+}(p_r + p_{r+})} , \end{aligned} \tag{4.12}$$

thus establishing the derivation of Pearson's expression for chi-square.

When sampling is done without replacement each variance and covariance is multiplied by $(N - n)/(N - 1)$ where $N$ is the number of observations in the universe. Hence, chi-square for this case can be written

$$\chi^2 = \frac{N - 1}{N - n} \sum_{i=1}^{k+1} \frac{e_i^2}{np_i} . \tag{4.13}$$

This expression shows that the factor involving sampling errors is the same whether sampling is done with replacement or without replacement. Hence, the derivation of least squares statistics is the same for either method of sampling, but sampling variances for the simpler case are multiplied by the factor $(N - n)/(N - 1)$ when sampling is done without replacement.

**5. The method of minimum chi-square.** The derivation of Pearson's expression for chi-square completes first four steps of the ideal method of least squares outlined in section 3. Hence, the method of minimum chi-square is the sample-frequency form of the ideal method of least squares in which only two of the six steps remain to be taken.

In his original article [4] Pearson pointed out that the use of statistics instead of parameters would affect the value of chi-square but that such effects would usually be so small that no allowance need be made for them in connection with tests of significance. It is now well known that the average value of chi-square is reduced approx:mately one unit for each parameter estimated from the sample, and that the main portion of this effect is on the numerators; i.e., in large samples the effect of substituting statistics, for parameters in the denominators usually has a negligible effect on the value of chi-square. By confining the discussion to the case in which parameters are used in the denominators, it is possible to make simple exact statements concerning the main effects in terms of the number of squares of standardized uncorrelated linear functions—also known as the number of degrees of freedom and the mean value of chi-square.

When the expected values in the numerators of chi-square can be expressed as linear functions of $r$ algebraically independent parameters, ideal linear estimates of the $r$ parameters are determined by substituting statistics for the $r$ parameters and minimizing the resulting expression wth respect to each statistic. In general, such a substitution of statistics for parameters in the numerators of chi-square reduces the number of degrees of freedom by one unit for every parameter estimated; that is, the appropriately minimized chi-square can be analyzed into $k - r$ squares of standardized uncorrelated linear functions of sampling errors.

The $r$ ideal linear estimates are linear functions of the sample frequencies. Let $(v_1, v_2, \cdots, v_r)$ be a set of standardized uncorrelated linear functions of the correlated sampling errors in these statistics and let $(v_1, v_2, \cdots, v_k)$ be a set of linear functions obtained from the $z_i$'s of section 3 by an orthogonal transformation. Since the sum of squares is not changed by such a transformation, chi-square is the sum of the $k$ values of $v_i^2$. The process of substituting statistics for the $r$ parameters in the numerators of chi-square reduces the values of the first $rv_i^2$'s to zero without affecting the values of the other $(k - r)v_i^2$'s.

Thus the appropriately minimized chi-square can be analyzed into $k - r$ squares of standardized uncorrelated linear functions of sampling errors and is therefore said to have $k - r$ degrees of freedom. The mean value of each square is the variance of a standardized linear function of sampling errors and is therefore unity by definition. Hence the mean value of the appropriately minimized chi-square (with parameters in the denominators) is exactly $k - r$ when $r$ statistics are estimated from a set of $k + 1$ sample frequencies.

The expression to be minimized is

$$\chi^2 = \sum \frac{(n_i - m_i')^2}{np_i} \tag{5.1}$$

where $m_i'$ is the ideal linear estimate of $np_i$. The set of statistics described by the equation

$$m_i' = n_i, \tag{5.2}$$

reduces the value of chi-square to zero—its minimum value. This shows that the sample cell proportion is the ideal linear estimate of the corresponding parameter.

Whenever a linear function independent of the sum of the cell proportions is known, it is possible to take advantage of additional information provided by the known function by minimizing chi-square subject to an appropriate side condition. When side conditions are used in this way, the number of degrees of freedom for the minimized chi-square is equal to the number of side conditions which are algebraically independent of each other (and of the sum of the cell proportions). Let the known linear function be written

$$\Sigma a_i np_i - m = 0. \tag{5.3}$$

In order to facilitate comparison of the typical equation of maximization with the corresponding equation of the method of maximum likelihood, it is convenient to minimize chi-square by maximizing $-\chi^2/2$ subject to a side condition based on (5.3). The function to be maximized can be written

$$-\chi^2/2 = \Sigma(n_i - m_i')^2/(-2np_i) + h(\Sigma a_i m_i' - m), \tag{5.4}$$

where $h$ is a Lagrange multiplier. Setting the partial derivative of $-\chi^2/2$ with respect to $m_i'$ equal to zero, the typical equation for minimizing chi-square can be written

$$(n_i - m_i')/np_i + ha_i = 0, \tag{5.5}$$

a form which shows that, in general, ideal linear estimates are defined in terms of unknown parameters. Fortunately, these parameters can usually be approximated closely by an iterative process. Substituting $m_i$ for both $np_i$ and $m_i'$ in equations (5.5) the typical equation in the limiting values of such a process can be reduced to

$$n_i/m_i - 1 + ha_i = 0, \tag{5.6}$$

a form which is identical with the typical equation (6.6) of maximum likelihood derived in section 6. This equality of typical equations implies that whenever the denominators of chi-square are estimated in such a way as to be consistent with least squares statistics based on them, the method of minimum chi-square always leads (by means of approximations necessary in practice) to maximum likelihood estimates of parameters which are linear functions of cell proportions.

**6. The method of maximum likelihood.** Maximum likelihood estimates of linear functions of cell proportions can be obtained by (1) expressing the probability function (general term of the multinomial expansion) in terms of the $r$ parameters to be estimated; (2) substituting $r$ statistics for the $r$ parameters; and (3) maximizing with respect to the $r$ statistics. In practice, this is usually accomplished by maximizing the logarithm of the variable factor in step (3) which can be written,

$$L = \Sigma n_i \log m_i, \tag{6.1}$$

where $m_i$ is the maximum likelihood estimate of $np_i$, the expected value of the $i$th frequency $n_i$ in a sample of $n$ observations classified into $(k + 1)$ classes or cells. It is evident that $L$ as written has no maximum with respect to any $m_i$ since it increases without bound as $m_i$ increases, but it sometimes has a uniquely determined maximum when each of the $m_i$'s is written explicitly in terms of less than $k + 1$ algebraically independent statistics. In the general case it is easier to maximize $L$ subject to an appropriate set of side conditions, one of which must be equivalent to

$$m_1 + m_2 + \cdots + m_{k+1} - n = 0. \tag{6.2}$$

When no linear function except the sum is known, the likelihood function can be written

$$L = \Sigma n_i \log m_i - (\Sigma m_i - n), \tag{6.3}$$

a function which, subject to equation (6.2), is always equal to that in equation (6.1) but which has a uniquely determined maximum. The typical equation of maximum likelihood, obtained by setting the partial derivative of $L$ with respect to $m_i$ equal to zero, is

$$n_i/m_i - 1 = 0, \tag{6.4}$$

an equation which shows that each sample frequency is a maximum likelihood estimate of its own expected value.

When a linear function such as that in equation (5.3) is known, an improved set of maximum likelihood statistics can be found by maximizing

$$L = \Sigma n_i \log m_i - (\Sigma m_i - n) + h(\Sigma a_i m_i - m). \tag{6.5}$$

The typical equation of maximization is found to be

$$n_i/m_i - 1 + ha_i = 0, \tag{6.6}$$

an equation which, as stated above, is identical with equation (5.5). Since equation (5.5) was obtained as the limit of an iterative process from the typical equation (5.4) for minimizing chi-square subject to the same side condition and since each additional side condition affects the typical equation of each method in exactly the same way, the method of minimum chi-square and the method of maximum likelihood are equivalent for the general case in the sense that the method of minimum chi-square always leads to maximum likelihood statistics as limits of an iterative process.

**7. Second-order tables with known expected marginal totals.** As stated in section 2, the integration of available techniques is facilitated by regarding maximum likelihood statistics as the best practical approximations to the corresponding ideal linear estimates. Since this important principle may not be immediately obvious, it will be illustrated for the important special case of second-order tables for which the expected marginal totals are known.

Consider a sample of $n$ observations arranged on two bases of classification and presented in a table containing $r$ rows and $s$ columns. The universe of $N$

observations has been completely enumerated and classified on each basis separately but not cross-classified; i.e., universe totals of first order classes are known.

For the cell in the $i$th row and the $j$th column, let $p_{ij}$ represent the universe cell proportion; $n_{ij}$, the sample frequency; $np_{ij}$, the expected value of $n_{ij}$; and $m_{ij}$, the maximum likelihood estimate of $np_{ij}$. Indicating summation by substituting a dot for the letter over which summation is to be performed, the known marginal totals satisfy the equations

$$\begin{aligned} (7.1) \qquad Np_{i.} - N_{i.} &= 0, \\ Np_{.j} - N_{.j} &= 0, \end{aligned}$$

where $p_{i.}$ and $p_{.j}$ are the universe proportions and $N_{i.}$ and $N_{.j}$ are the known universe totals in the $i$th row and the $j$th column, respectively.

When $n$ observations of a random sample are arranged according to two bases of classification in a table with $r$ rows and $s$ columns for which the $r + s$ marginal totals are known, the typical equation of maximum likelihood can be obtained by maximizing, subject to side conditions based on equations (7.1), the likelihood function

$$(7.2) \qquad L = \Sigma\Sigma n_{ij} \log m_{ij} - \Sigma a_i(m_{i.} - n_{i.}) - \Sigma b_j(m_{.j} - n_{.j}),$$

with respect to the maximum likelihood estimates $m_{ij}$, where $a_i$ and $b_j$ are typical Lagrange multipliers. Setting the partial derivative with respect to $m_{ij}$ equal to zero and transposing, the typical equation of maximum likelihood can be written

$$(7.3) \qquad n_{ij}/m_{ij} = a_i + b_j .$$

Since equations (7.3) are not linear in their unknowns, the reader's first reaction might well be to agree with a certain anonymous critic that "their solution is difficult." This impression of great difficulty is probably the chief reason that previous writers have not used the method of maximum likelihood for this type of problem even after they had developed a set of techniques adequate for the solution of the equations of maximum likelihood. In other words, all that was needed was the integration of available techniques as will now be shown.

In 1940, Deming and Stephan [2] derived a set of normal equations for the adjustment of a set of second-order cell frequencies to known expected marginal totals by the method of least squares in which each sample frequency is weighted by its own reciprocal. This method yields statistics which are efficient according to the theory of large samples, but they do not satisfy the criterion of maximum likelihood exactly. In the same article was presented an easier method of *iterative proportions*, which, unfortunately, does not yield least squares statistics. In 1942, Stephan [5] developed an improved iterative process which yields statistics which satisfy the criterion of least squares with arbitrarily chosen weights. The foregoing developments are presented in greater detail in Deming's book [1] in which Deming adapts Stephan's iterative method to the particular case in which each sample frequency is weighted by its own reciprocal so as to yield solutions for the normal equations derived in the joint article [2].

In Deming's notation, equation 8 of Stephan's article [5, p. 169] can be written

$$m_{ij} = c_{ij}(p_i + q_j - 1) + n_{ij}, \tag{7.4}$$

an expression obtained by substituting $c_{ij}$ for $np_{ij}$ in the denominators of chi-square and minimizing with respect to the statistics in the numerators. Hence, if exact values of the $np_{ij}$ were used for the $c_{ij}$, the Stephan iterative method would yield ideal linear estimates. Unless these parameters are implied by some hypothesis to be tested, it is necessary, in practice, to estimate the $np_{ij}$ from sample data. In order to secure maximum likelihood estimates of expected cell frequencies by means of the Stephan iterative method, the adjusted frequencies based on first approximations to the $c_{ij}$ should be used as second approximations to the $c_{ij}$, etc. In this way, maximum likelihood statistics can be derived to any desired degree of approximation. At this point it should be emphasized that the preceding statement applies not only to the class of problems considered in this section but also to the wider class of problems for which the Stephan iterative method provides solutions.

Unfortunately, theoretical discussions of previous writers contain confusing compensating errors which (1) present their own methods in an unnecessarily unfavorable light and (2) increase the difficulties involved in the introduction of the improvements in techniques suggested in section 9 which involve some degree of adaptation of techniques already available. For these reasons, it seems necessary to follow the arguments of previous writers in order to show the points at which improvements are needed. This can be done most effectively in connection with Deming's book [1] where the method of least squares is presented in great detail.

For the special case in which the sampling errors in the observations are uncorrelated, the ideal criterion of least squares implies that the weight of each observation should be inversely proportional to its sampling variance. This criterion is accepted as well known by Deming who says that "the principle of least squares requires the minimizing of the sum of the weighted squares of the residuals" [1, p. 14] where "the weights of two functions are inversely proportional to their variances" [1, p. 22]. Deming assumes that "there is no correlation between the errors in the observations" with the qualification that "this assumption covers a wide class of problems, but does fail to cover some." [1, p. 49]. This assumption of uncorrelated errors is not applicable to sample-frequency problems, of course, because the sample frequencies are correlated with each other in such a way that the reciprocals of the ideal least squares weights are not proportional to the sampling variances $np_{ij}q_{ij}$ but rather to the expected frequencies $np_{ij}$ which appear in the denominators of chi-square.

In this connection it is interesting to note that Deming himself insists that "there is only one principle of least squares, namely, the minimizing of $\chi^2$." [1, p. 51]. However, the method currently in use for the minimizing of chi-square was that given by Fisher [3] which leads to equations which are difficult to solve even for such a simple example as the one presented in section 2 above.

Deming and Stephan are to be commended for seeking an easier method but there is no justification (even as a device for saving effort) for their modification of the "principle of least squares" so as to imply erroneously that

(1) weights of correlated sample frequencies are inversely proportional to their variances, and

(2) sample frequencies are, in general, approximately proportional to their own sampling variances.

Strangely enough, these two errors were applied in combination by Deming and Stephan to obtain good practical approximations to the ideal least squares weights. It might be argued that the second misleading implication is really not an error because it is offered as a simplifying approximation, but it is an integral part of both the normal equations approach in the joint article [2] and Deming's adaptation [1] of the Stephan iterative method; that is, in each case the method would have to be revised if better approximations to the ideal least squares weights were used. More explicitly, Deming (1) uses $n_{ij}$ for Stephan's $c_{ij}$ in equation (7.4); (2) identifies it with the other $n_{ij}$ in the same equation; and (3) reduces the equation to a different form thus effectively preventing the use of successive approximations to the $c_{ij}$ without returning to Stephan's iterative method in the general form given by equation (7.4) above which Deming does not present at all. Results of the joint article [2] are quoted by Stephan [5] without any explanation of the nature of the errors, but none of these results are used in the development of his iterative method which as noted above, is applicable to any arbitrarily chosen set of weights. The fact that Stephan corrected the second error without correcting the first implies that the weights he actually used are unsatisfactory. In Deming's adaptation of the Stephan iterative method, a much better set of weights is obtained, not by correcting the first offsetting error overlooked by Stephan, but by resurrecting the second offsetting error which Stephan had corrected. Since this error is an integral part of Deming's adaptation, Deming's theoretical discussion implies that his own efficient statistics are only rough approximations which are definitely inferior to the inefficient statistics obtained by means of the weights chosen by Stephan. These inconsistencies are most clearly brought out by Deming when he says:

> "Strictly, in random sampling, the reciprocal of the weight of $n_{ij}$ is $np_{ij}q_{ij}$, which is nearly equal to $n_{ij}q_{ij}$ where $p$ and $q$ have their usual connotations. But since factors proportional to the weights may be substituted for them, it is sufficient to use $n_{ij}$ as the reciprocal of the weight in cell $ij$, since the values of $q_{ij}$ do not usually vary much over the table." [1, p. 102.]

In any given problem the seriousness of the error in the first statement in the foregoing quotation depends on the variation among the $q_{ij}$'s. In the particular example used by Deming the error is of considerable importance because the largest $q_{ij}$ is more than 40 per cent larger than the smallest $q_{ij}$. The weights actually used by Deming agree with weights implied by the ideal method of least squares except for sampling errors in the $n_{ij}$; hence, the error in any relative weight converges stochastically to zero so that Deming's statistics are efficient according to the theory of large samples. The efficiency of Deming's statistics is inconsistent with the theory presented by Deming which implies erroneously that efficiency of estimation depends on approximate equality of cell proportions. If this argument were true it would apply also to the method of maximum likelihood and all other methods which yield efficient practical statistics in sample-frequency problems. The foregoing discussion, together with the results of section 8 show that the theory as presented by Deming has the following seriously misleading features:

(1) it is based on a paradox in which a good final result is obtained by means of compensating errors;

(2) it presents his efficient statistics in an unnecessarily unfavorable light;

(3) it emphasizes the irrelevant condition of approximate equality of universe cell proportions;

(4) it fails to mention the important condition of proportionality by rows and columns; and

(5) it makes least squares, minimum chi-square, and maximum likelihood seem to be competing alternative methods.

Of these undesirable characteristics, the last two are probably the most serious because they make the effective integration and adaptation of statistical techniques more difficult. As has been shown in sections 4, 5, and 6, the sample-frequency form of the ideal method of least squares is the method of minimum chi-square which always leads (by means of appropriate practical approximations to unknown weights) to maximum likelihood statistics; in other words, the methods are equivalent from a practical point of view.

Since the ideal method of least squares based on the unknown $np_{ij}$ determines fully efficient, but theoretical, ideal linear estimates, the efficiency of practical approximations to ideal linear estimates depends on the accuracy with which the denominators of chi-square are estimated. For the unknown denominators $np_{ij}$, Deming uses the sample frequencies $n_{ij}$ while the method of maximum likelihood implies the use of the corresponding maximum likelihood estimates—statistics which, in general, have smaller sampling variances. The foregoing argument suggests that maximum likelihood statistics are slightly superior to Deming's statistics for any given finite value of $n$ and that their relative advantage increases as the sample size decreases. In large samples both methods yield efficient statistics because the relative errors in the weights implied by either method converge stochastically to zero as $n$ increases. Although the advantage of maximum likelihood statistics over Deming's statistics is unimportant except in small samples, it can be shown that Deming's choice of weights leads to imperfectly compensated negative errors of estimation even in his large sample of 33,837 observations.

Deming weights each sample frequency by its own reciprocal. Positive errors of sampling decrease the value of the reciprocal and thus increase the absolute size of the required negative adjustments. Negative errors of sampling increase the value of the reciprocal and thus decrease the size of the positive adjustment. Thus every error of sampling (either positive or negative) leads to a negative error of estimation due to inappropriate weighting. Because the sum of all adjustments must be zero, these negative errors of estimation are compensated on the average but more or less imperfectly. The net effect of this imperfect compensation of negative errors of estimation is that Deming's statistics are too small in those cells in which the relative adjustments (either positive or negative) are large, and vice versa. In a preliminary draft of this article, this type of error of estimation was studied by comparing Deming's statistics with the corresponding maximum likelihood statistics in conection with Deming's example involving 33,837 observations. Although errors of estimation of the type under discussion are apparent, they are, of course, extremely small in such a large sample. For this reason the large-sample comparson has been deleted in favor of simple hypothetical examples designed to throw light on similar errors of estimation in statistics derived by Fisher's method of minimum chi-square as well as in those derived by Deming's adaptation of Stephan's iterative method.

Consider a set of sample frequencies in a two-by-two table for which all expected marginal totals are equal. For this special case, the cell proportions on each diagonal are equal and the ideal linear estimate (which is also the maximum likelihood estimate) of any cell proportion is the mean of the two sample cell proportions on its diagonal. For the same case, Deming's adaptation of the Stephan iterative method yields an estimate for each cell which is proportional to the harmonic mean of sample proportions on its diagonal while Fisher's method of minimum chi-square yields estimates proportional to the corresponding quadratic means.

As a numerical example of the foregoing problem consider the set of frequencies

$$n_{11}, n_{12}, n_{21}, n_{22} = 1, 4, 3, 2, \tag{7.5}$$

obtained in a sample of 10 observations selected at random from a universe in which the cell poportions are known to be

$$p_{11}, p_{12}, p_{21}, p_{22} = p, 0.5 - p, 0.5 - p, p. \tag{7.6}$$

As estimates of the parameter $p$, the ideal linear estimate is .15, Deming's adaptation of the Stephan iterative method yields .14, and Fisher's method of minimum chi-square yields .1545 to four decimal places, the other two estimates being exact. The results illustrate the imperfectly compensated errors of estimation explained previously. The two sample frequencies on the principal diagonal ($n_{11}$ and $n_{22}$) have greater relative dispersion than the frequencies on the other diagonal. For this reason, the relative adjustments made by Deming's method are greater and according to the principle of imperfectly compensated negative errors of estimation, the estimate of $p$ obtained by Deming's method is smaller than the ideal linear estimate of $p$. Fisher's method of minimum chi-square yields an estimate of $p$ which is greater than the ideal linear estimate. In fact, one should usually expect imperfectly compensated errors of estimation in statistics derived by Fisher's method of minimum chi-square to be opposite in sign and about half as large as those in the corresponding statistics derived by means of Deming's adaptation of the Stephan iterative method.

At this point, it should be emphasized that Fisher does not recommend his own method of minimum chi-square in preference to the method of maximum likelihood. In fact, he presents the theory of estimation in such a way as to imply correctly that the method of maximum likelihood is superior, especially in small samples. Other writers have noted the small differences between equations of maximum likelihood and those for minimizing chi-square by Fisher's method and some have even derived one set of equations from the other by neglecting higher order terms in a Taylor series expansion. These derivations are of no interest here because they seem to justify the method of maximum likelihood as a simple approximation to some more complicated method. This type of justification is both unnecessary and undesirable. It is more useful to regard the method of maximum likelihood as an approximation to a method—least squares—for which the theory is simpler.

Skeptical readers who find the foregoing argument unconvincing may be able to profit from the following example. Consider the problem of estimating the parameter $p$ where $2p$ is the proportion of white balls in an urn. A sample of 10 balls is selected and classified by the following process. Each white ball is placed in one of the cells on the principal diagonal of a two-by-two table, the particular cell being decided by the toss of a coin. A similar methcd is used for non-white balls placed in cells on the other diagonal. Assuming that the results of this process are given by equation (7.5), which of the three alternative estimates of $p$ given above should be preferred? Belief in the general superiority of Fisher's method of minimum chi-square seems to imply that the device of coin-tossing described in this example can be used in practical problems involving the estimation of the proportion of "successes" to secure estimates which are superior to the sample proportion—the ideal linear estimate in such cases. Even if it is possible to construct trivial special case examples supporting some complicated method for such problems the general use in practical problems of the coin-tossing device in connection with either Fisher's method of minimum chi-square or Deming's adaptation of the Stephan iterative method would be absurd as this example is intended to emphasize.

**8. The method of proportional distribution of marginal adjustments.** The method of proportional distribution of marginal adjustments is a general method of adjusting sample frequencies so that their row and column totals agree with known expected marginal totals. In other words, the adjusted frequency for the cell in the $i$th row and the $j$th column is given by the equation

$$(8.1) \qquad m^{*}_{ij} = n_{ij} - p_{i.}d_{.j} - p_{.j}\, d_{i.}\,,$$

where

$$d_{i.} = m_{i.} - n_{i.}\,,$$

and

$$d_{.j} = m_{.j} - n_{.j}\,,$$

are the net adjustments in the sample cell frequencies of the $i$th row and the $j$th column, respectively. The asterisk is used to distinguish maximum likelihood estimates $m_{ij}$ and the ideal linear estimates $m'_{ij}$ from the set of statistics based on equation (8.1).

The method of proportional distribution of marginal adjustments yields ideal linear estimates when the universe cell proportions are proportional by rows and by columns; i.e., when

$$(8.2) \qquad p_{ij} = p_{i.}p_{.j}\,.$$

This important principle can be established by substituting in equation (7.4) of section 7 the quantities

$$(8.3) \qquad c_{ij} = np_{i.}p_{.j}\,,$$

$$p_i = 0.5 + d_{i.}/np_{i.}\,,$$

and

$$q_j = 0.5 + d_{.j}/np_{.j}\,,$$

and reducing the typical equation of the ideal method of minimum chi-square to the form of equation (8.1) which defines the method of proportional distribution of marginal adjustments.

Even in the absence of exact proportionality, under which it yields fully efficient statistics, the method of proportional distribution of marginal adjustments has the following relative advantages over other available methods:

(1) ease of extension to tables of higher order;
(2) *exact* agreement with known (expected) marginal totals;
(3) simplicity of interpretation;
(4) independence of computational errors;
(5) rapidity of processing;
(6) economy of effort; and
(7) fully efficient criteria for testing the significance of departures from proportionality of rows and columns.

Ease of extension to tables of higher order is a desirable property of the method of proportional distribution of marginal adjustments. Equation (8.1) applies to the special case in which there are only two bases of classification. In the more general case sample observations are cross-classified according to $r$ bases of classification, each cell frequency in an $r$th order table being the number of observations in the corresponding $r$th order class whose expected value is to be estimated. The required adjustment for each first order class (obtained by subtracting the sample total from its known expected value) is distributed among the various cells in proportion to the universe totals of the corresponding $(r - 1)$th order classes to which the cells belong. The general process is illustrated by

$$m^*_{ijk} = n_{ijk} + p_{i..}d_{.jk} + p_{.j.}d_{i.k} + p_{..k}d_{ij.}, \tag{8.4}$$

the formula for estimating the expected frequency in the general cell of a third order table.

Exact agreement with marginal totals follows easily from the method of proportional distribution and can be established algebraically by summing the estimation equation by first order classes; e.g., summing equation (8.1) by rows and columns. In practice, discrepancies are always either errors of rounding or mistakes in computation; they are never due to lack of convergence of iterative processes as is often true in alternative methods of estimation.

Although simplicity of interpretation is desirable in general, it is especially important when random sampling is an unrealistic abstraction. For example, the method of proportional distribution of marginal adjustments has been used to estimate the cell proportions in a two-way classification of incomes from known marginal proportions and a detailed cross classification at an earlier date. In this problem known shifts in income distributions made it evident that certain cells previously vacant should not have the zero proportions which would be estimated for them by other available methods of estimation. The ease with which the effects of the method of adjustment can be traced is important also in the analysis of the results of sample surveys in which various types of bias are important.

The method of proportional distribution of marginal adjustments yields the estimated expected frequency for any cell by a single sequence of computations which is independent of the corresponding process for any other cell. Errors made in computing the estimate for any cell appear in marginal totals of estimates for all first order classes which include that cell. If only a few errors are made in a table they can be localized immediately and can be corrected without recomputing any estimates which are correct.

In certain types of social surveys, rapidity of processing is so important that, as Deming puts it, "the delay of only the brief time required for adjustment may not be advisable." [1, p. 102]. Under these conditions, it is important to have a simple formula like equation (8.1) in which substitutions can be made rapidly. Even when the time element is relatively unimportant, the economy of effort and the ease of explaining the method to clerical assistants are often of practical importance.

Finally, departures from proportionality among rows and columns often provide the chief element of interest in research studies—not only in social surveys of the type illustrated in Deming and Stephan's example but also in biological sciences. The most effective tests of significance for the purpose of presenting statistical evidence of lack of proportionality are those based on statistics like those derived by the method of proportional distribution of marginal adjustments whose efficiency is 100 per cent when proportionality is exact.

Even when proportionality is not exact, the efficiency of statistics derived by proportional distribution may be close to 100 per cent under fairly typical problem conditions such as those in the example by Deming and Stephan wherein the other more complicated methods require several times as much computational effort, but have little advantage over the easier method with respect to efficiency of estimation in this particular problem.

**9. Suggested improvements in techniques.** In section 7, a method was outlined by which it is possible to derive sets of maximum likelihood statistics by merely integrating available techniques without changing any of them. In this section a number of improvements are suggested. At this point it should be emphasized that a given change is not an improvement merely because it yields slightly more accurate estimates or makes possible a slight saving of time and effort. In each case the research worker should consider saving of time and effort and accuracy of estimation simultaneously. In particular, it seems likely that most social surveys of the type considered by Deming and Stephan are characterized by approximate proportionality by rows and by columns—conditions relatively favorable to the simple method of proportional distribution of marginal adjustments. It should be clearly understood that suggestions in this section are intended for those research workers whose problems justify a great deal more effort than is required to adjust sample frequencies by this simple method.

Assuming that the problem at hand warrants the effort required to derive maximum likelihood estimates, the first consideration is the derivation of a set of $m_{ij}(1)$, first approximations to the $m_{ij}$, and a set of values of $p_i(1)$, first approximations to the $p_i$. Even if proportionality by rows and by columns is not closely approximated use of values of the $p_i(1)$ provided by equation (8.3) are especially to be recommended. In the example used by Deming these values for the $p_i(1)$ are so much better than the values recommended by Deming that they save a large proportion of the effort required by the iterative process. If rows and columns are approximately proportional, equation (8.1) should be used to provide values of the $m_{ij}(1)$, in which case it is possible to use an iterative process similar to the one used by Deming but based on the typical equation of maximum likelihood (7.3) to achieve a given degree of accuracy in the maximum likelihood estimates with even less effort. Under favorable conditions such as those in Deming's example the suggested iterative process yields excellent approximations to maximum likelihood estimates by means of the following steps:

1. Construct a set of first approximations to the $r$ row components of the $rs$ maximum likelihood divisors $(a_i + b_j)$ by means of the equation

$$a_i(1) = n_{i.}/np_{i.} - 1/2. \tag{9.1}$$

2. Compute successive approximations to the $a_i$ and $b_j$ by means of the equations

$$b_j(g) = [n_{.j} - \Sigma m_{ij}(1)a_i(g)]/np_{.j}, \tag{9.2}$$

$$a_i(g + 1) = [n_{i.} - \Sigma m_{ij}(1)b_j(g)]/np_{i.}, \tag{9.3}$$

where $m_{ij}(1)$, the first approximation to $m_{ij}$, is derived by means of equation (8.1). Just as in Deming's iterative process, the expression in brackets is a series of products which can be subtracted in a single sequence of machine operations and the final division can be performed without having to record any of the intermediate results.

3. Divide the sample frequencies by the maximum likelihood divisors to obtain the maximum likelihood estimates

$$m_{ij} = n_{ij}/(a_i + b_j), \tag{9.4}$$

where limiting values of $a_i$ and $b_j$ are approximated as closely as desired by successive approximations in the preceding equations.

Under unfavorable conditions, the iterative process of this section is not always the easiest way to obtain satisfactory estimates. For example, when samples are small and/or rows and columns are not approximately proportional, it is better to use the iterative method as originally presented by Stephan where sample frequencies can be used for first approximations to the $c_{ij}$ and these may be replaced by successively better approximations.

The point made in the final paragraph of Fisher's well-known book [3] that "in practice one need seldom do more than solve, at least to a good approximation, the equation of maximum likelihood," is strongly supported by the developments of this article. In addition, the proof that the method of least squares and the method of minimum chi-square always lead (by means of approximations to ideal weights) to maximum likelihood statistics greatly facilitates the adaptation of techniques developed in connection with these hitherto competing methods.

## REFERENCES

[1] W. Edwards Deming, *Statistical Adjustment of Data*, John Wiley & Sons, 1943.

[2] W. Edwards Deming and Frederick F. Stephan, "On a least squares adjustment of a sample frequency table when the expected marginal totals are known," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 427–444.

[3] R. A. Fisher, *Statistical Methods for Research Workers*, 6th ed., Oliver and Boyd, 1936, Ch. 9.

[4] Karl Pearson, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *Phil. Mag.*, Vol. 50 (1900), pp. 157–175.

[5] Frederick F. Stephan, "An iterative method of adjusting sample frequency tables when expected marginal totals are known," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 166–178.

[6] E. T. Whittaker, and G. Robinson, *The Calculus of Observations*, D. Van Nostrand Company, 1924, Ch. 9.

# A STATISTICAL PROBLEM CONNECTED WITH THE COUNTING OF RADIOACTIVE PARTICLES

By Sten Malmquist

*Institute of Statistics, University of Upsala, Sweden*

**1. Introduction.** Our problem refers to random events forming a sequence in time or in space, *e.g.* particles emitted by a radioactive matter. By omitting certain elements of the given sequence, say $f$, we form another sequence, say $g$. The rule of omission involves an arbitrarily prescribed constant $u$. The rule to be followed in forming $g$ is:

*Case I*: Let $a$ be an element in $f$ and $g$. The next element to be included in $g$ is then the first element in $f$ which follows $a$ after a distance greater than $u$.

*Case II*: Let $a$ be an element in $f$ and $g$. The next element to be included in $g$ is then the first element in $f$ which follows $a$ at a distance greater than $u$ from the preceding element in $f$, whether this belongs to $g$ or not.

When the events are represented by impulses emitted by a radioactive matter and feeding a recorder with a constant resolving time $u$, the new sequence consists of the counted impulses. The two cases correspond to the reaction of different types of recorders. The distinction between the two transformations has caused some confusion. It has, however, been clearly pointed out by Ruark and Brammer [5].

v. Bortkiewicz [2] seems to be the first who has considered problems related to the transformed sequence. Starting from investigations by Rutherford, Geiger, and others, concerning the number of recorded $\alpha$-particles during a certain interval of time, say $T$, he observed that the distribution of this number was similar to that of Poisson but with a slightly smaller dispersion. This fact he supposed to be caused by a constant resolving time $u$ of the recorder. By means of certain assumptions he tried to calculate the effect on the mean and the dispersion by the transformation in Case I, supposing the cumulative distribution function $F(t)$ for the distance between two consecutive elements in the sequence $f$ is given by

$$F(t) = 1 - e^{-at},$$

where here and in what follows, $t$ denotes a non-negative variable.

Considering Case II with $F(t)$ as above, Levert and Scheen [4] have recently worked out an expression for the distribution of the number of elements during $T$ in the sequence $g$.

Gnedenko [3] has considered the distribution of the number of lost elements in Case I with particular regard to the initial state of rest.

Alaoglu and Smith [1] considered problems referring to successive transformations of a sequence. When, for example, a sequence of particles enters a tube-counter and amplifier, together acting with a resolving time $u_1$, and the impulses then are feeding a recorder with resolving time $u_2 > u_1$, the sequence of recorded impulses will be the result of two successive transformations. If we have a scaling circuit between the counter and the recorder, we have to make a transformation of another type between the two transformations in Case I and Case II.

The present paper deals with the transformed sequence in Case I. The distribution function $F(t)$ is supposed to be arbitrary. An advantage of this generalization is that the formulas derived could be used in treating problems referring to successive transformations.

The author wishes to express his sincere gratitude to Professor Herman Wold for stimulating discussions and valuable advice.

**2. Derivation of distributions for case I.** Suppose that the sequence $f$ has $F(t)$ for distribution function for the distance between two consecutive elements. $F(t)$ is supposed to be independent of absolute time (space), and of the preceding distance between two elements. When not stated otherwise, we further suppose $F(0) = 0$.

Now let $G(t)$ be the distribution function for the distance between two consecutive elements in the transformed sequence $g$. Evidently $G(t)$ also is independent of absolute time and of the preceding distance between two elements.

We shall consider certain distribution functions connected with $F(t)$. These functions will then be used in solving problems concerning the sequence $g$.

Let $F_n(t)$ be the distribution function for the distance between the first and the last of $n + 1$ consecutive elements in the sequence $f$. Then $F_n(t)$ is given by the recursive system

$$(1)\qquad \begin{aligned} F_{m+n}(t) &= \int_0^t F_m(t - x)\, dF_n(x); \qquad (m, n \geqq 1) \\ F_1(t) &\equiv F(t). \end{aligned}$$

As is easily seen, we have

$$F_{m+n}(t) \leq F_m(t) \cdot F_n(t);$$

and, for $t = u$,

$$F_n(u) \to 0, \qquad \text{as } n \to \infty;$$

$$\sum_{n=1}^{\infty} F_n(u) < \infty, \qquad \text{provided that } F_1(0) < 1.$$

Alternatively, $F_n(t)$ could be deduced by the use of characteristic functions.

Still considering the sequence $f$, let $\Phi(t)$ be the distribution function for the distance $d$ between an arbitrarily chosen point and the following element. Suppose that the arbitrary point is chosen so that the distance between the preceding and the following element is $x$. Under this condition we have, in usual symbols,

$$P(d > t) = \frac{x - t}{x}.$$

Hence,

$$\Phi(t) = 1 - \int_t^\infty \frac{x - t}{x}\, dH(x)$$

where $H(t)$ is the distribution function for the distance $x$.

To deduce $H(t)$ we suppose that the distribution $F(t)$ has a finite mean,

$$m = \int_0^\infty t\, dF(t).$$

By the definition of $H(t)$, we then have

$$H(x) = \frac{1}{m}\int_0^x tdF(t).$$

Thus

$$\Phi(t) = \frac{1}{m}\left[\int_0^t x\, dF(x) + t\int_t^\infty dF(x)\right]. \tag{2}$$

The corresponding frequency function $\varphi(t)$ is given by

$$\varphi(t) = \frac{1 - F(t)}{m}.$$

Consider $n + 2$ consecutive elements in $f$, say $a_0, a_1, \cdots, a_{n+1}$, where $a_0$ is an element in the transformed sequence $g$. The probability $P_n$ that the next element in $g$ following $a_0$ will be $\mathrm{a}_{n+1}$ is given by

$$P_n = F_n(u) - F_{n+1}(u), \qquad (n = 1, 2, \cdots),$$
$$P_0 = 1 - F(u).$$

Now let $P_n(t)$ be the probability that the distance between $a_0$ and $a_{n+1}$ is smaller than or equal to $t$, when $a_0$ an $a_{n+1}$ are two consecutive elements in the sequence $g$. Then

$$P_n(t) = \frac{1}{F_n(u) - F_{n+1}(u)} \int_0^u \Big[F(t - x) - F(u - x)\Big]\, dF_n(x),$$
$$(n = 1, 2 \cdots), \qquad P_0(t) = \frac{F(t) - F(u)}{1 - F(u)}.$$

Let $G^*(t)$ be defined by

$$G^*(t) = \sum_{n=0}^\infty P_n \cdot P_n(t) = F(t) - F(u)$$
$$+ \sum_{n=1}^\infty \int_0^u [F(t - x) - F(u - x)]\, dF_n(x); \qquad t > u.$$

When $G^*(t)$ is a distribution function, then $G^*(t)$ equals $G(t)$.

For $t_1 < t_2$ we obviously have $G^*(t_1) \leq G^*(t_2)$.

For $t = \infty$

$$\begin{aligned} G^*(\infty) &= 1 - F(u) + \sum_{n=1}^{\infty} \int_0^u [1 - F(u - x)]\, dF_n(x) \\ &= 1 - F(u) + \sum_1^{\infty} F_n(u) - \sum_1^{\infty} F_{n+1}(u) = 1. \end{aligned}$$

Hence we take

$$\begin{aligned} G(t) &= G^*(t); & t &> u, \\ G(t) &= 0; & t &\leq u. \end{aligned} \tag{4}$$

When the corresponding frequency functions $g(t)$ and $f(t)$ exist, we get

$$g(t) = f(t) + \sum_{n=1}^{\infty} \int_0^u f(t - x) f_n(x)\, dx; \qquad t > u. \tag{5}$$

Dealing with a sequence of elements we are often concerned with the number of occurrences during a certain time $T$.

Let the mean number of occurrences during $T$ be $M(T)$. Supposing that the mean $m = \int_0^{\infty} t\, dF(t)$ is finite and that $F(0) < 1$, we have

$$M(T) = T/m. \tag{6}$$

We define

$$K_1(t) = \begin{matrix} F(t) & \text{for } t \geq \epsilon \\ 0 & \text{for } t < \epsilon \end{matrix}$$

$$K_2(t) = \begin{matrix} F(t) & \text{for } t \geq \epsilon \\ F(\epsilon) & \text{for } t < \epsilon \end{matrix}$$

and denote the corresponding means by $M_1(T)$ and $M_2(T)$. As is easily seen,

$$M_1(\epsilon) \leq M(\epsilon) \leq M_2(\epsilon).$$

Using (2),

$$M_1(\epsilon) = \frac{\epsilon F(\epsilon) + \epsilon[1 - F(\epsilon)]}{\int_0^{\infty} x\, dK_1(x)} = \frac{\epsilon}{\int_0^{\infty} x\, dK_1(x)};$$

$$\begin{aligned} M_2(\epsilon) &= \frac{1}{\int_0^{\infty} x\, dK_2(x)} [1 \cdot \epsilon[1 - F(\epsilon)]^2 + \cdots + n \cdot \epsilon[1 - F(\epsilon)]^2 F(\epsilon)^{n-1} + \cdots] \\ &= \frac{\epsilon}{\int_0^{\infty} x\, dK_2(x)}. \end{aligned}$$

Making $N = T/\epsilon$ and summing, we obtain

$$M_1(T) = \frac{T}{\displaystyle\int_0^\infty x\, dK_1(x)} = \frac{T}{m - \displaystyle\int_0^\epsilon x\, dF(x) + \epsilon F(\epsilon)};$$

$$M_2(T) = \frac{T}{\displaystyle\int_0^\infty x\, dK_2(x)} = \frac{T}{m - \displaystyle\int_0^\epsilon x\, dF(x)}.$$

By choosing $\epsilon$ arbitrarily small, we get

$$M(T) \rightarrow T/m.$$

Let $P(n, T)$ be the probability that we get $n$ elements in $f$ during a time $T$. Suppose that the first of these elements, $a_1$, comes at $T_0 + x$, and the last, $a_n$, at $T_0 + x + y$.

We then have

$$P(n, T) = \int_0^T \varphi(x)\, dx \int_0^{T-x} [1 - F(T - x - y)]\, dF_{n-1}(y). \tag{7}$$

In (4) and (7) we have equations for the transformation in Case I. Because of the general form of $F(t)$, the formulas also can be used when we are concerned with successive transformations. It can further be remarked that the transformation of a sequence of impulses by passing a scaling circuit is expressed by the system (1).

**3. Results for a particular form for** $F\ (t)$. The preceding formulas will now be used for a special distribution function $F(t)$. Suppose that the frequency function $f(t) = dF(t)/dt$ is equal to the frequency function of the distance between an arbitrary point and the following element.

From (3) we get

$$F'(t) = \frac{1 - F(t)}{m},$$

or, when $F(0) = 0$,

$$F(t) = 1 - e^{-at}; \tag{8}$$

$$f(t) = ae^{-at}, \quad \text{where } 1/a = m = \int_0^\infty tf(t)\, dt. \tag{9}$$

By means of the theory of characteristic functions we have

$$f_n(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} [\eta(x)]^n e^{-itx}\, dx; \qquad f_1(t) \equiv f(t); \tag{10}$$

where

$$\eta(x) = a \int_0^\infty e^{-at} e^{itx}\, dt = \frac{a}{a - ix}. \tag{11}$$

Thus

$$f_n(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{a^n}{(a - ix)^n}\, e^{-itx}\, dx \tag{12}$$

For $n = 1$, we get

$$f_1(t) = ae^{-at} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{a}{a - ix}\, e^{-itx}\, dx \tag{13}$$

By differentiating (13) $n - 1$ times with respect to $a$ we obtain

$$(-t)^{n-1} e^{-at} = \frac{1}{2\pi}(-1)^{n-1}(n-1)! \int_{-\infty}^{+\infty} \frac{e^{-itx}}{(a - ix)^n}\, dx.$$

Hence, from (12),

$$f_n(t) = \frac{a^n}{(n-1)!}\, t^{n-1} e^{-at}. \tag{14}$$

From (5) we obtain the frequency function for the transformed sequence $g$

$$\begin{gathered} g(t) = ae^{-at} + \sum_{n=1}^{\infty}\int_0^u ae^{-at}\,\frac{a^n}{(n-1)!}\, t^{n-1}\, dx = ae^{au} e^{-at}; \qquad t \geq u \\ G(t) = 0;\ t < u. \end{gathered} \tag{15}$$

The mean $m_g$ is given by

$$m_g = a\int_u^{\infty} te^{au} e^{-at}\, dt = \frac{1}{a} + u.$$

*Remark:* Suppose the constant $u$ is allowed to vary independently of $t$ and that the frequency function of $u$ is $\gamma(u)$, we obtain

$$\begin{aligned} m_g &= \int_0^{\infty} t\, dt \int_0^t g(u, t)\gamma(u)\, du = \int_0^{\infty} \frac{1}{a}\,\gamma(u)\, du + \int_0^{\infty} u\gamma(u)\, du \\ &= \frac{1}{a} + m(u). \end{aligned} \tag{16}$$

Now let the sequence of elements, $g$, by means of (5) be transformed into a new sequence, $h$. When we are concerned with the counting of particles, emitted from a radioactive matter, let the sequence $g$ consist of impulses from a counter-amplifier with resolving time $u$, feeding a recorder with resolving time $u_1$. Then the elements in $h$ are the counted impulses, it being supposed that the tube-counter and the recorder reacts according to the assumptions.

We suppose $u_1 > u$. When $u_1 \leq u$, the sequences $g$ and $h$ are identical.

Let $g_n(t)$ denote the frequency function of the distance between the first and the last of $n + 1$ consecutive elements in $g$. We find, in the same way as used in obtaining (14),

$$g_n(t) = \frac{a^n}{(n-1)!}\, e^{anu}(t - nu)^{n-1} e^{-at}; \qquad t \geq nu. \tag{17}$$

Let $h(t)$ be the frequency function for the distance between two consecutive elements in the sequence $h$. Let further $N$ be the greatest integer smaller than or equal to $u_1/u$.

Using (4) and (5) we obtain

$$\begin{aligned} h_I(t) &= ae^{au}e^{-at}\sum_0^N \frac{a^n}{n!}(u_1 - nu)^n e^{anu}; && t \geq u_1 + u; \\ (18) \qquad h_{II}(t) &= ae^{au}e^{-at}\sum_0^N \frac{a^n}{n!}[t - (n+1)u]^n e^{anu}, && (N+1)u \leq t \leq u_1 + u; \\ h_{III}(t) &= ae^{au}e^{-at}\sum_0^{N-1} \frac{a^n}{n!}[t - (n+1)u]^n e^{anu}, && u_1 \leq t \leq (N+1)u. \end{aligned}$$

The mean $m_h$ is found to be

$$(19) \qquad m_h = \left[\frac{1}{a} + u\right]\left[1 + \sum_{n=1}^{N}\sum_{v=n}^{\infty}\frac{(u_1 - nu)^v a^v}{v!} e^{-a(u_1 - nu)}\right].$$

We also have

$$\int_{u_1+u}^{\infty} th_I(t)\,dt < m_h < \int_u^{\infty} th_I(t)\,dt$$

or

$$\left[\frac{1}{a} + u_1 + u\right]\left[\sum_0^N \frac{a^n}{n!}(u_1 - nu)^n e^{-a(u_1 - nu)}\right] < m_h < \left[\frac{1}{a} + u_1\right]e^{au}\left[\sum_0^N \frac{a^n}{n!}(u_1 - nu)^n e^{-a(u_1 - nu)}\right].$$

We now consider the number of occurrences during a time interval $T$. Using (6), (16), and (19) we immediately get the mean numbers of occurrences during $T$.

By (3), we get for the sequence $g$

$$(20) \qquad \varphi_g(t) = \begin{cases} \dfrac{a}{au + 1}; & t \leq u \\ \dfrac{a}{au + 1}e^{au}e^{-at}; & t \geq u. \end{cases}$$

Inserting (20), (15) and (14) in (7) and evaluating the integrals, we finally get

$$(21) \qquad P_g(n, T) = \begin{cases} a_{n-1} - 2a_n + a_{n+1}; & n \leq \dfrac{T}{u} - 1 \\ a_{n-1} - 2a_n + (n+1) - \dfrac{aT}{au + 1}; & \dfrac{T}{u} - 1 \leq n \leq \dfrac{T}{u} \\ a_{n-1} - 2\left[n - \dfrac{aT}{au + 1}\right] + (n+1) - \dfrac{aT}{au + 1}; & \\ & \dfrac{T}{u} \leq n \leq \dfrac{T}{u} + 1. \end{cases}$$

where

$$a_n = \frac{1}{au+1}\, e^{-a(T-nu)} \sum_{v=0}^{n} \frac{(T - nu)^v\, a^v}{v!}\,(n - v), \qquad (n = 0, 1, \cdots), \tag{22}$$

$$a_{-1} = 0.$$

When $u = 0$, we obtain

$$a_n = e^{-aT} \sum_{v=0}^{n} \frac{T^v\, a^v}{v!}\,(n - v).$$

For the sequence $f$ we then get the Poisson distribution

$$P_f(n, T) = \frac{(aT)^n}{n!}\, e^{-aT}. \tag{23}$$

The corresponding expression for the sequence $h$ is much more complicated.

**4. A statistical experiment.** The following statistical experiment will serve as an illustration of the scheme dealt with in this paper—the transformation of a sequence and the resulting formulas, especially (21).

Groups of five figures, the last rounded up if necessary, have been extracted from tables of random sampling numbers (6). Let each group denote the first five digits for a decimal $x$, arbitrarily chosen between 0 and 1. The variable $x$ is supposed to have the distribution function $t$ for $0 \leq t < 1$. We now define a new variable, $y$, given by

$$y = -k \log (1 - x), \text{ [or } y = -k \log x]. \tag{24}$$

The variable $y$ has the distribution function given by (8), *viz.*

$$F(t) = 1 - e^{-at}, \text{ where } \frac{1}{a} = m = k \log e.$$

Transforming each group, or number $x$, according to (24), we get a sample of consecutive distances between elements in the sequence $f$ considered in the previous sections. Choosing a constant $u$, we can construct the corresponding sequence $g$. Beginning with a point, arbitrarily chosen on the first distance, we can finally count the number of elements in successive intervals of the same length.

Take $k = 1$, $u = 0.2$ and $T = 1.5$. We then have for the sequences $f$ and $g$:

$$m_f = \frac{1}{a} = \log e = 0.4343; \qquad m_g = \frac{1}{a} + u = 0.6343;$$

$$\sigma_f = \frac{1}{a} = 0.4343; \qquad \sigma_g = \frac{1}{a} = 0.4343;$$

$$M_f(T) = \frac{T}{m_f} = 3.454. \qquad M_g(T) = \frac{T}{m_g} = 2.365.$$

The experiment yielded the following results:

| *For the sequence f:* | *For the sequence g:* |
|---|---|
| Number of elements 801. | Number of elements 555. |
| $\bar{m}_f = 0.450.$ | $\bar{m}_g = 0.648.$ |

In neither case is the deviation between the observed and theoretical means statistically significant. In fact we have:

$$\frac{(\bar{m}_f - m_f)\sqrt{800}}{\sigma_f} \sim 1.0; \qquad \frac{(\bar{m}_g - m_g)\sqrt{554}}{\sigma_g} \sim 0.8$$

which gives $P = 0.3$ and $P = 0.4$, respectively.

TABLE I

*Nos. of intervals with n elements*

| $n$ | Sequence $f$ | | Sequence $g$ | | |
|---|---|---|---|---|---|
| | Observed | Expected according to (23) | Observed | Expected according to (21) | Expected according to (23) |
| 0 | 6 | 7.6 | 5 | 8.2 | 23.7 |
| 1 | 33 | 26.1 | 53 | 42.5 | 54.8 |
| 2 | 48 | 45.1 | 82 | 81.8 | 63.3 |
| 3 | 55 | 51.9 | 69 | 72.2 | 48.8 |
| 4 | 36 | 44.8 | 23 | 29.2 | 28.1 |
| 5 | 32 | 31.0 | 6 | 4.8 | 13.0 |
| 6 | 17 | 17.8 | 1 | 0.2 | 5.0 |
| 7– | 12 | 14.7 | | | 2.4 |
| $\Sigma$ | 239 | 239 | 239 | 238.9 | 239 |
| Mean | 3.331 | 3.454 | 2.310 | 2.36 | 2.31 |
| $\chi^2$ | | 4.825 | | 4.524 | 36.7 |
| $P$ | | 0.68 | | 0.34 | <0.001 |

The functions $a_n$ in (22) can be calculated by means of Pearson's tables of the incomplete $\gamma$-function (7). In the notation of these tables we obtain

$$e^{-\lambda}\sum_{v=r}^{\infty}\frac{\lambda^v}{v!} = I\left(\frac{\lambda}{\sqrt{r-1}}\,; r-2\right) = I(p, q).$$

Hence

$$a_n = \frac{n}{au+1}\,e^{-\lambda}\frac{\lambda^n}{n!} + \frac{n-\lambda}{au+1}\,[1 - I(p, q)],$$

where

$$\lambda = a(T - nu); \qquad p = \frac{\lambda}{\sqrt{n - 1}}; \qquad q = n - 2.$$

In the present case, however, we only need the numbers up to $a_7$. Accordingly, the $a_n$ have been calculated directly.

The resulting theoretical and observed distributions for the number of elements during $T$ for the sequences $f$ and $g$ will be found in Table I. For comparison, a Poisson distribution, with the same mean as observed for the sequence $g_1$ is given. The result of a $\chi^2$ test is also shown in Table I. Judged by the $\chi^2$ test the distributions (23) and (21) agree fairly well with the observed distributions. As was to be expected, the Poisson distribution cannot be used for the sequence $g$.

## REFERENCES

[1] L. Alaoglu and N. M. Smith, "Statistical theory of a scaling circuit," *Phys. Rev.*, Vol. 53 (1938), pp. 832–836.

[2] L. v. Bortkiewicz, *Die radioaktive Strahlung als Gegenstand wahrscheinlichkeitstheoretischer Untersuchungen*, Berlin, 1913.

[3] B. V. Gnedenko, "On the theory of Geiger-Müller counters," (in Russian), *Jour. for Exp. and Theor. Phys.*, Vol. 11 (1941), pp. 101–106.

[4] C. Levert and W. L. Scheen, "Probability fluctuations of discharges in a Geiger-Müller counter produced by cosmic radiation," *Physica*, Vol. 10 (1943), pp. 225–238.

[5] A. E. Ruark and F. E. Brammer, "The efficiency of counters and counter circuits," *Phys. Rev.*, Vol. 52 (1937), pp. 322–324.

[6] M. G. Kendall and B. Babington Smith, *Tables of Random Sampling Numbers*, Tracts for Computers XXIV, Cambridge, 1939.

[7] Karl Pearson, *Tables of the Incomplete* Γ*-function*, Cambridge, 1922.

# THE PROBABILITY FUNCTION OF THE PRODUCT OF TWO NORMALLY DISTRIBUTED VARIABLES[1]

By Leo A. Aroian

*Hunter College*

**1. Introduction and summary.** Let $x$ and $y$ follow a normal bivariate probability function with means $\bar{X}$, $\bar{Y}$, standard deviations $\sigma_1$, $\sigma_2$, respectively, $r$ the coefficient of correlation, and $\rho_1 = \bar{X}/\sigma_1$, $\rho_2 = \bar{Y}/\sigma_2$. Professor C. C. Craig [1] has found the probability function of $z = xy/\sigma_1\sigma_2$ in closed form as the difference of two integrals. For purposes of numerical computation he has expanded this result in an infinite series involving powers of $z$, $\rho_1$, $\rho_2$, and Bessel functions of a certain type; in addition, he has determined the moments, seminvariants, and the moment generating function of $z$. However for $\rho_1$ and $\rho_2$ large, as Craig points out, the series expansion converges very slowly. Even for $\rho_1$ and $\rho_2$ as small as 2, the expansion is unwieldy. We shall show that as $\rho_1$ and $\rho_2 \to \infty$, the probability function of $z$ approaches a normal curve and in case $r = 0$ the Type III function and the Gram-Charlier Type A series are excellent approximations to the $z$ distribution in the proper region. Numerical integration provides a substitute for the infinite series wherever the exact values of the probability function of $z$ are needed. Some extensions of the main theorem are given in section 5 and a practical problem involving the probability function of $z$ is solved.

**2. Theorems on approach to normality.** The moment generating function of $z$, $M_z(\theta)$, is [1]

$$(2.1)\qquad M_z(\theta) = \frac{\exp \dfrac{(\rho_1^2 + \rho_2^2 - 2r\rho_1\rho_2)\theta^2 + 2\rho_1\rho_2\theta}{2[1-(1+r)\theta][1+(1-r)\theta]}}{\sqrt{[1-(1+r)\theta][1+(1-r)\theta]}}.$$

Let $\bar{z}$, and $\sigma_z$ be the mean and the standard deviation of $z$, and $t_z = (z - \bar{z})/\sigma_z$. Now

$$(2.2)\qquad \bar{z} = \rho_1\rho_2 + r, \qquad \sigma_z = \sqrt{\rho_1^2 + \rho_2^2 + 2r\rho_1\rho_2 + 1 + r^2}.$$

Using (2.2) we find in the usual way the moment generating function of $t_z$

$$(2.3)\quad M_{t_z} = \frac{\exp \dfrac{-2rw + (\rho_1^2 + \rho_2^2 + 2r\rho_1\rho_2)w^2 + 4r^2w^2 - 2w^3(r^2-1)(\rho_1\rho_2 + r)}{2[1-(1+r)w][1+(1-r)w]}}{\sqrt{[1-(1+r)w][1+(1-r)w]}},$$

where $w = \theta/\sigma_z$.

[1] Presented to the American Mathematical Society, Oct. 28, 1944, New York City.

Consider $r \geqq 0$. Then in the limit as $\rho_1$ and $\rho_2 \to \infty$ in any manner whatever,

$$\lim_{\rho_1, \rho_2 \to \infty} M_{t_z}(\theta) = e^{\theta^2/2}, \tag{2.4}$$

and by the theorem of Curtiss [2] on moment generating functions we see in the limit as $\rho_1, \rho_2 \to \infty$ the probability function of $z$ approaches a normal curve with mean, $\bar{z}$, and variance $\sigma_z^2$, $r \geqq 0$.

In case $-1 + \epsilon < r < 0$, $\epsilon > 0$, some care is required wherever

$$\sqrt{\rho_1^2 + \rho_2^2 + 2\rho_1\rho_2 r}$$

occurs. If one uses $\rho_1^2 + \rho_2^2 \geqq 2\rho_1\rho_2$, the proof goes forward quite readily. Hence we have proved the theorem:

THEOREM (2.5). *The distribution of $z$ approaches normality with mean $\bar{z}$, and variance $\sigma_z^2$ as $\rho_1$ and $\rho_2 \to \infty$ in any manner whatever, $-1 + \epsilon < r \leqq 1$, $\epsilon > 0$.*

It is evident in Theorem (2.5) we may allow $\rho_1, \rho_2 \to -\infty$ without any other changes. Theorems (2.6) and (2.7) are proved in essentially the same way as (2.5).

THEOREM (2.6). *The distribution of $z$ approaches normality with mean $\bar{z}$, and variance $\sigma_z^2$, if $\rho_1 \to \infty$, $\rho_2 \to -\infty$, $-1 \leqq r < 1 - \epsilon$, $\epsilon > 0$.*

THEOREM (2.7). *The distribution of $z$ approaches normality if $\rho_1$ remains constant $\rho_2 \to \infty$, $-1 + \epsilon < r \leqq 1$, $\epsilon > 0$; or if $\rho_1$ remains constant $\rho_2 \to -\infty$, $-1 \leqq r < 1 - \epsilon$, $\epsilon > 0$.*

Naturally in any of the theorems $\rho_1$ and $\rho_2$ may be interchanged. In practice $\rho_1$ and $\rho_2$ are usually positive. The approach to normality is more rapid if both $\rho_1$ and $\rho_2$ have the same sign as $r$.

**3. Numerical values.** In order to show how closely the Type III and the Gram-Charlier Type A series approximate the probability function of $z$, $f(z)$, or more precisely $f(z, \rho_1, \rho_2, r)$, we use numerical integration where

$$f(z, \rho_1, \rho_2, r) = I_1(z) - I_2(z),$$

$$I_1(z) = \frac{1}{2\pi\sqrt{1-r^2}} \int_0^\infty \exp - \frac{1}{2(1-r^2)} \left\{ (x-\rho_1)^2 - 2r(x-\rho_1)\left(\frac{z}{x} - \rho_2\right) + \left(\frac{z}{x} - \rho_2\right)^2 \right\} \frac{dx}{x}, \tag{3.1}$$

and $I_2(z)$ is the integral of the same function over $(-\infty, 0)$, [1]. Now $I_1(z)$ may be written as

$$I_1(z) = \frac{1}{\sqrt{1-r^2}} \int_0^\infty \varphi(t_1)\varphi(t_2)\beta(t_3)\frac{dx}{x}, \tag{3.2}$$

where

$$\varphi(t) = \frac{e^{-(t^2/2)}}{\sqrt{2\pi}}, \qquad t_1 = \frac{x-\rho_1}{\sqrt{1-r^2}}, \qquad t_2 = \left(\rho_2 - \frac{z}{x}\right) \Big/ \sqrt{1-r^2},$$

$$\beta(t_3) = e^{t_3}, \; t_3 = rt_1t_2.$$

We readily obtain $I_1(z)\sqrt{1 - r^2}$ by forming the product of $\varphi(t_1)$, $\varphi(t_2)$, $\beta(t_3)$, and $1/x$ using numerical integration applying Weddle's formula, the Gregory-Newton formula, or the simple rectangular formula depending on circumstances. The rectangular formula [3] is remarkably accurate when the function $T = \varphi(t_1)\varphi(t_2)\beta(t_3)/x$ in the interval 0 to $\infty$ or 0 to $-\infty$ is somewhat symmetrical. Appropriate tables for $\varphi(t_1)$, $\varphi(t_2)$ (see [4]), $\beta(t_3)$ (see [5]) and $1/x$ (see [6]) are readily available. In the important case of the independence of $x$ and $y$, $r = 0$ and (3.2) becomes

$$(3.3) \qquad I_1(z) = \int_0^\infty \varphi(t_1)\varphi(t_2)\,\frac{dx}{x}, \qquad t_1 = x - \rho_1, \qquad t_2 = \rho_2 - \frac{z}{x}.$$

**4. Approximations to $f(z)$.** When $r = 0$, the standard seminvariants $\xi_3$, and $\xi_4$ of $z$ are

$$(4.1) \qquad \xi_3 = \frac{6\rho_1\rho_2}{(\rho_1^2 + \rho_2^2 + 1)^{3/2}}, \qquad \xi_4 = \frac{6\{2(\rho_1^2 + \rho_2^2) + 1\}}{(\rho_1^2 + \rho_2^2 + 1)^2} \quad .$$

remembering

$$\bar{z} = \rho_1\rho_2\,, \ \sigma_z = \sqrt{\rho_1^2 + \rho_2^2 + 1}.$$

In the Pearson system (see [7]) $\delta$, the criterion, is

$$(4.2) \qquad \delta = \frac{2\xi_4 - 3\xi_3^2}{6 + \xi_4}$$

and for the probability function of $z$

$$(4.3) \qquad \delta = \frac{2(\rho_1^2 + \rho_2^2 + 1)\{2(\rho_1^2 + \rho_2^2) + 1\} - 18\rho_1^2\rho_2^2}{(\rho_1^2 + \rho_2^2 + 1)[(\rho_1^2 + \rho_2^2 + 1)^2 + 2(\rho_1^2 + \rho_2^2) + 1]}$$

and if $\rho_1 = \rho_2 = \rho$

$$(4.4) \qquad \delta = \frac{2(4\rho^2 + 1)(2\rho^2 + 1) - 18\rho^4}{(2\rho^2 + 1)[(2\rho^2 + 1)^2 + (4\rho^2 + 1)]}.$$

Now $\delta = 0$, $\xi_3 \neq 0$, for the Type III function, and clearly $\lim_{\rho_1,\rho_2\to\infty} \delta = 0$. By use of (3.3) the accurate values of $f(z)$ have been calculated for various combinations of $\rho_1$ and $\rho_2$ and compared with the Type III approximation using $\bar{z}$, $\sigma_z$, $\xi_3$.

(4.5) Investigations so far completed show that for $\rho_1 \geqq 4$ and $\rho_2 \geqq 4$ simultaneously, and $|\delta| \leqq .008$, the Type III approximation will provide values of $t_z$ correct to three significant figures at least where

$$(4.6) \qquad \int_{-\infty}^{t_z^{(1)}} f(t_z) = \alpha, \qquad \int_{t_z^{(2)}}^{\infty} f(t_z) = \alpha, \quad \text{and} \quad .05 \leqq \alpha \leqq .005.$$

These are the values of $t_z$ which would be needed in testing hypotheses. The exact values of $t_z^{(1)}$ and for $t_z^{(2)}$ for various values of $\rho_1$ and $\rho_2$ less than 4 will be determined it is hoped in the future and will be published along with the comparisons of the Type III values of $t_z$ with the accurate values of $t_z$ in the important borderline cases of $\rho_1 = \rho_2 = 2$, and $\rho_1 = \rho_2 = 3$. The values of $f(z)$ for $\rho_1 = \rho_2 = 2$ and $\rho_1 = \rho_2 = 4$ have been calculated but these are being withheld for a more complete table. The table of values of $\bar{z}$, $\sigma_z$, $\xi_3$, $\xi_4$, and $\delta$ (Table II) shows then that the Type III function is excellent along a band about $\rho_1 = \rho_2$, since $\xi_3 \neq 0$, and $\delta$ is very small.

We use the Gram-Charlier Type A series of three terms to approximate the probability function of $z$ in $t_z$ units.

$$f(t_z) \sim \varphi(t) - \frac{\xi_{3:z}}{3!}\varphi^{(3)}(t) + \frac{\xi_{4:z}}{4!}\varphi^{(4)}(t), \tag{4.7}$$

in the usual notation.

TABLE I

| $t_z$ | $f(t_z)$ Correct value | Normal Curve | Gram-Charlier Type A |
|---|---|---|---|
| .9950372 | .2406367 | .2431716 | .2408235 |
| 1.4925558 | .1275209 | .130970 | .127484 |
| 1.9900744 | .0538243 | .0550708 | .053704 |
| 2.4875930 | .0184606 | .0180791 | .0184500 |
| 2.9851116 | .0052477 | .0046338 | .0052944 |
| 3.4826302 | .0012609 | .0009272 | .0012804 |
| 3.9801488 | .0002611 | .0001449 | .000260 |
| 4.4776674 | .0000467 | .0000177 | .0000425 |
| 4.9751860 | .00000745 | .00000168 | .00000555 |

(4.8) For $|\xi_3| < .5$ and $\xi_4 < .4$ simultaneously the Gram-Charlier Type A series is quite adequate for finding probability levels such as those of (4.6). These will in general give 3 significant figures for $t_z^{(1)}$ or $t_z^{(2)}$. In the special case $\rho_1 = 0$, $\rho_2 = 10$, the Gram-Charlier Type A series differs from $f(t_z)$ very slightly in the range $1 \leqq |t_z| < \infty$ (see Table 1). Naturally the Gram-Charlier will be used wherever Type III is not indicated, although there exist some overlapping regions where either one may be used. It should be noticed that the approach of $f(z)$ to normality is more rapid along a row than down a diagonal. In case either $\rho_1$ or $\rho_2$ is negative, we may make use of the equation

$$f(z, -\rho_1, \rho_2, r) = f(-z, \rho_1, \rho_2, -r). \tag{4.9}$$

We note that when $r = 0$, $f(z, \rho_1, \rho_2)$ always possesses a discontinuity at $z = 0$, (see [1]). A table of $\bar{z}$, $\sigma_z$, $\xi_3$, $\xi_4$, and $\delta$ is provided for values of $\rho_1$ and $\rho_2$ from 0 to 10 inclusive.

TABLE II*

| $\rho_2$ \ $\rho_1$ | 2 | 4 | 6 | 8 | 10 |
|---|---|---|---|---|---|
| 0 | 0<br>2.236068<br>0<br>2.160<br>.529 | 0<br>4.123106<br>0<br>.685121<br>.205 | 0<br>6.082762<br>0<br>.319942<br>.101 | 0<br>8.062258<br>0<br>.183195<br>.059 | 0<br>10.049876<br>0<br>.118224<br>.039 |
| 2 | 4<br>3<br>.8<br>1.259259<br>.020 | 8<br>4.582576<br>.498784<br>.557823<br>.056 | 12.<br>6.403124<br>.274256<br>.289114<br>.056 | 16.<br>8.306624<br>.167493<br>.172653<br>.042 | 20.<br>10.246951<br>.111531<br>.113742<br>.031 |
| 4 | | 16.<br>5.744563<br>.506408<br>.358127<br>−.0084 | 24.<br>7.280110<br>.373206<br>.224279<br>.0049 | 32.<br>9.<br>.263374<br>.147234<br>.014 | 40<br>10.816654<br>.189641<br>.102126<br>.016 |
| 6 | | | 36.<br>8.544004<br>.346314<br>.163258<br>−.0054 | 48.<br>10.049876<br>.28373<br>.118224<br>−.00083 | 60<br>11.704700<br>.224503<br>.087272<br>.0038 |
| 8 | | | | 64.<br>11.357817<br>.262088<br>.092663<br>−.0034 | 80<br>12.845233<br>.226472<br>.072507<br>−.0015 |
| 10 | | | | | 100.<br>14.177447<br>.210551<br>.059553<br>−.0023 |

* The first value in *a* cell is $\bar{z}$, the second $\sigma_3$, the third $\xi_3$, the fourth $\xi_4$, the fifth $\delta$.

**5. Some extensions.** We may generalize our results to any case where $x$ and $y$ are distributed approximately in a normal distribution such as the distribution of the product of two means, when the sizes of the samples $N_1$ and $N_2$ are large and consequently $\rho_1$ and $\rho_2$ will be large. Another example occurs if $x$ and $y$ each follows a Bernoulloi probability function with parameters $p_1$ and $p_2$ respectively where the number of trials in each case is large. We must warn the reader that the condition $\rho_1 \rightarrow \infty$, $\rho_2 \rightarrow \infty$ alone does not mean that the distribution of $z$ approaches normality. Both $x$ and $y$ must be distributed normally.

The actual problem which gave rise to this investigation was the question of determining the sum of a great many variates [8]. Let $T$ variates $v_1, v_2, \cdots, v_T$ be given whose sum $A = \sum_{i=1}^{T} v_i$ is desired. Clearly

$$A = T\bar{V}_p, \quad \bar{V}_p = \sum_{i=1}^{T} v_i/T.$$

Now let us estimate $A$ by $\tilde{A} = \tilde{T}_s\bar{V}_s$ where $\tilde{T}_s$ is an estimate of $T$ and $\bar{V}_s$ is an estimate of $\bar{V}_p$. If $\sigma_{\tilde{T}_s}$ is very small, $\rho_1 = T/\sigma_{\tilde{T}_s}$ will be large and $\rho_2 = \bar{V}_p/\sigma_{\bar{V}_s} = \sqrt{N}\bar{V}_p/\sigma_p$ will be very large. Assuming $\tilde{T}_s$ is distributed normally and obviously $\bar{V}_s$ is distributed normally for $N$ large, we see by the theorems of this paper that $\tilde{A}$ will be distributed normally. Confidence limits for $A$ may be calculated in the usual fashion as $\tilde{A} \pm \gamma\sigma_{\tilde{A}}$, where $\gamma$ is determined by

$$\int_{t=\gamma}^{\infty} \varphi(t)dt = \alpha,$$

with $\alpha$ generally chosen as .025 or less and

$$\sigma_{\tilde{A}} = \sqrt{\tilde{T}_s^2\sigma_{\bar{V}_s}^2 + \bar{V}_s^2\sigma_{\tilde{T}_s}^2 + \sigma_{\bar{V}_s}^2\sigma_{\tilde{T}_s}^2}.$$

Stratification is also possible. It is interesting to note that many functions which occur in life insurance are products. Such applications will be treated fully elsewhere. Naturally the critical region whether both tails or one tail of the distribution should be used depends on the alternatives to the hypothesis being tested.

Generalizations of the main theorem are possible for the probability function of $z = \prod_{i=1}^{r} x_i$ where $x_1, x_2, \cdots, x_r$ follow a multivariate normal probability function. These will be investigated in a later paper. It may be noted that J. B. S. Haldane has investigated the distribution of a product along different lines [9].

## REFERENCES

[1] CECIL C. CRAIG, "On the frequency function of $xy$," *Annals of Math. Stat.*, Vol. 7 (1936), pp. 1–15.

[2] J. H. CURTISS, "A note on the theory of moment generating functions," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 430–434.

[3] A. L. O'TOOLE, "On the degree of approximation of certain quadrature formulas," *Annals of Math. Stat.*, Vol. 4 (1933), pp. 143–153.

[4] ARNOLD N. LOWAN, Technical Director, *Tables of Probability Functions, Vol. II*. National Bureau of Standards, Washington, D. C.

[5] ARNOLD N. LOWAN, Technical Director, *Tables of the Exponential Function $e^x$*. National Bureau of Standards, Washington, D. C.

[6] ARNOLD N. LOWAN, Technical Director, *Tables of Reciprocals of Integers from 100,000 through 200,009*. Columbia Univ. Press, New York.

[7] CECIL C. CRAIG, "A new exposition and chart for the Pearson system of frequency curves," *Annals of Math. Stat.*, Vol. 7 (1936), pp. 16-28.

[8] LEO A. AROIAN, "Some methods for the evaluation of a sum," *Amer. Stat. Ass. Jour.*, Vol. 39 (1944), pp. 511-515.

[9] J. B. S. HALDANE, "Moments of the distribution of powers and products of normal variates," *Biometrika*, Vol. 32 (1942), pp. 226-242.

# NOTES

*This section is devoted to brief research and expository articles on methodology and other short items.*

## A REMARK ON CHARACTERISTIC FUNCTIONS

By A. Zygmund

*University of Pennsylvania*

**1.** Let $F(x)$, $-\infty < x < +\infty$, be a distribution function, and

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx}\, dF(x)$$

its characteristic function. It is well known that the existence of $\varphi'(0)$ does not imply the existence of the absolute moment

$$\int_{-\infty}^{+\infty} | x |\, dF(x). \tag{1}$$

A simple example is provided by the function

$$\varphi(t) = C \sum_{n=2}^{\infty} \frac{\cos nt}{n^2 \log n},$$

where $C$ is a positive constant. Since the series on the right differentiated term by term converges uniformly (see [1]), $\varphi'(t)$ exists (and is continuous) for all values of $t$, and in particular at the point $t = 0$. Obviously $\varphi(t)$ is the characteristic function of the masses $C/2n^2 \log n$ concentrated at the points $\pm n$ for $n = 2, 3, \cdots$. The constant $C$ is such that the sum of all the masses is 1. The divergence of the series $\Sigma 1/n \log n$ implies that in this particular case the moment (1) is infinite.

In a recent paper (see [2], esp. p. 120, footnote), Fortet raises the problem of whether the existence of $\varphi'(0)$ implies the existence of the first algebraic moment

$$\int_{-\infty}^{+\infty} x\, dF(x) = \lim_{X \to +\infty} \int_{-X}^{X} x\, dF(x). \tag{2}$$

The main purpose of this note is to show that this is so. We shall even prove a slightly more general result.

A function $\psi(t)$ defined in the neighborhood of a point $t_0$ is said to be *smooth* at this point if

$$\lim_{h \to +0} \frac{\psi(t_0 + h) + \psi(t_0 - h) - 2\psi(t_0)}{h} = 0.$$

Clearly, if $\psi$ has a one-sided derivative at the point $t_0$, the derivative on the other side also exists and has the same value. Thus the graph of $\psi(t)$ has no angular point for $t = t_0$, and this explains the terminology. If $\psi'(t_0)$ exists and is finite, $\psi(t)$ is smooth for $t = t_0$. The converse is obviously false, since any function whose graph is symmetric with respect to $t = t_0$ is smooth at that point.

THEOREM 1. *If the characteristic function $\varphi(t)$ is smooth at the point 0, then a necessary and sufficient condition for the existence of $\varphi'(0)$ is the existence of the moment* (2). *The value of* (2) *is* $-i\varphi'(0)$.

In particular, the existence and finiteness of $\varphi'(0)$ implies the existence of (2). That the converse is false, is obvious. For if $a_0, a_1, a_2, \cdots$ are positive numbers and $a_0 + 2a_1 + 2a_2 + \cdots = 1$, then $\psi(t) = a_0 + 2\Sigma_1^\infty a_n \cos nt$ is the characteristic function of the distribution function $F(x)$ corresponding to masses concentrated at the integer points $\pm n$ and having the values $a_n$ there. Owing to the symmetry of the masses, the number (2) exists, and is zero even if $\varphi(t)$ is non-differentiable for $t = 0$ (we may e.g. take for $\varphi(t)$ the Weierstrass non-differentiable function $C\,\Sigma_1^\infty a^n \cos b^n t$, where $C$ is a suitable constant).

PROOF. We may write

$$\varphi(t) = \int_0^\infty \cos xt\, dG(x) + i\int_0^\infty \sin xt\, dG(x) = \psi_1(t) + i\psi_2(t)$$

where

$$G(x) = F(x) - F(-x), \quad H(x) = F(x) + F(-x).$$

Thus

$$0 \le |\,\Delta H\,| \le \Delta G. \tag{3}$$

Since $\varphi(t)$ is smooth at the point 0, and since $\psi_1(t)$ is even, $\psi_2(t)$ odd,

$$0 = \lim_{h\to+0} \frac{\varphi(h) + \varphi(-h) - 2\varphi(0)}{h} = 2\lim \frac{\psi_1(h) - \psi_1(0)}{h}$$

$$= -2\lim_{h\to+0} \int_0^\infty \frac{1-\cos hx}{h}\, dG(x)$$

so that, replacing $h$ by $2h$,

$$\int_0^\infty \frac{\sin^2 hx}{h}\, dG(x) \to 0 \qquad \text{as } h \to 0.$$

Since the integrand is positive we obtain successively

$$\int_0^{1/h} \frac{\sin^2 hx}{h}\, dG(x) = o(1),$$

$$\int_0^{1/h} \frac{\left(\frac{2}{\pi}hx\right)^2}{h}\, dG(x) = o(1),$$

$$\int_0^{1/h} x^2\, dG(x) = o(h^{-1}), \tag{4}$$

$$\int_{1/2h}^{1/h} x^2\, dG(x) = o(h^{-1}),$$

$$\int_{1/2h}^{1/h} dG(x) = o(h). \tag{5}$$

Since $\psi_1(t)$ is even, the smoothness of $\varphi(t)$, and so also of $\psi_1(t)$, at the point $t = 0$ implies that $\psi_1'(0)$ exists and is zero. If $h \to +0$,

$$\frac{\psi_2(h) - \psi_2(0)}{h} = \int_0^\infty \frac{\sin xh}{h}\,dH(x) = \int_0^{1/h} + \int_{1/h}^\infty = A_h + B_h,$$

$$|B_h| \le h^{-1}\int_{1/h}^\infty |dH| \le h^{-1}\left(\int_{1/h}^{2/h} dG + \int_{2/h}^{4/h} dG + \int_{4/h}^{8/h} dG + \cdots\right)$$

$$= h^{-1}o(h + h/2 + h/4 + \cdots) = o(1),$$

by (3) and (5). Also

$$A_h - \int_0^{1/h} x\,dH = \int_0^{1/h}\left(\frac{\sin hx}{hx} - 1\right)x\,dH = \int_0^{1/h} O(x^2h^2)x\,dG$$

$$= \int_0^{1/h} O(x^2h)\,dG = o(1),$$

by (3) and (4). Thus

$$\frac{\psi_2(h) - \psi_2(0)}{h} = o(1) + \int_0^{1/h} x\,dH = o(1) + \int_{-1/h}^{1/h} x\,dF,$$

and so

$$\frac{\varphi(h) - \varphi(0)}{h} = o(1) + i\int_{-1/h}^{1/h} x\,dF.$$

It follows that the existence of (2) is equivalent to the existence of the right-hand side derivative of $\varphi(t)$ at the point $t = 0$, or, on account of smoothness, to the existence of $\varphi'(0)$. Moreover, the value of (2) is $-i\varphi'(0)$. This completes the proof of Theorem 1.

**2.** Suppose that a function $\psi(t)$ defined near the point $t_0$ satisfies for $h \to 0$ a relation

$$\psi(t_0 + h) = \alpha_0 + \alpha_1 h/1! + \cdots + \alpha_{k-1}h^{k-1}/(k-1)! + [\alpha_k + \sigma(1)]h^k/k!,$$

where $\alpha_0, \alpha_1, \cdots, \alpha_k$ are constants. Then $\alpha_k$ is called the *kth generalized derivative* of $\psi$ at the point $t_0$. It will be denoted by $\psi_{(k)}(t_0)$. The existence and finiteness of $\psi^{(k)}(t_0)$ implies the existence of $\psi_{(k)}(t_0)$ and both numbers are equal.

Another generalization of higher derivatives is based on the consideration of the symmetric differences

$$\Delta_h\psi(t_0) = \psi(t_0 + h) - \psi(t_0 - h),$$

$$\Delta_h^2\psi(t_0) = \psi(t_0 + 2h) - 2\psi(t_0) + \psi(t_0 - 2h),$$

$$\Delta_h^3\psi(t_0) = \psi(t_0 + 3h) - 3\psi(t_0 + h) + 3\psi(t_0 - h) - \psi(t_0 - 3h).$$

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

If $\Delta_h^k\psi(t_0)/(2h)^k$ tends to a limit as $h \to +0$, this limit is called the $k$th symmetric derivative of $\psi$ at the point $t_0$. We shall denote it by $D_k\psi(t_0)$. Clearly, $D_k\psi(t_0)$ exists and equals $\psi_{(k)}(t_0)$, if the latter number exists.

It is a simple matter to prove (see [3]) that if $k$ is a positive even integer, and if the characteristic function $\varphi(t)$ has at $t = 0$ a finite symmetric derivative $D_k\varphi(0)$, then the $k$th moment $\int_{-\infty}^{+\infty} x^k\, dF(x)$ exists, and its value is $(-1)^{k/2}D_k\varphi(0)$. Conversely, the existence of $\int_{-\infty}^{+\infty} x^k\, dF(x)$ obviously implies (for $k$ even) the existence and continuity of $\varphi^{(k)}(t)$ for all $t$, and in particular at the point $t = 0$.

In order to obtain an extension of Theorem 1 to the case of derivatives of odd order, we have to generalize the notion of smoothness. We shall say that a function $\psi(t)$ satisfies for $t = t_0$ condition $S_k$, $(k = 1, 2, \cdots)$, if

$$\Delta_h^{k+1}\psi(t_0) = o(h^k) \qquad \text{as } h \to +0.$$

For $k = 1$, condition $S_k$ is identical with smoothness at $t_0$. Clearly, if $\psi_{(k)}(t_0)$ exists, $\psi$ satisfies condition $S_k$ at $t_0$.

THEOREM 2. *Suppose that $k$ is a positive odd integer, and let $\varphi(t)$ be the characteristic function of a distribution function $F(x)$. If $\varphi$ satisfies condition $S_k$ at the point* 0, *a necessary and sufficient condition for the existence of $D_k\varphi(0)$ is the existence of the symmetric moment*

$$(6) \qquad \int_{-\infty}^{\infty} x^k\, dF(x) = \lim_{X\to+\infty} \int_{-X}^{X} x^k\, dF(x)$$

*whose value is then equal to $i^{-k}D_k\varphi(0)$. In particular, the existence of $\varphi_{(k)}(0)$ implies that of* (6).

The proof of Theorem 2 is analogous to that of Theorem 1. Let $G(x)$ and $H(x)$ have the same meaning as before. Since $k + 1$ is even, condition $S_k$ at the point $t = 0$ gives

$$\Delta_h^{k+1}\varphi(0) = \int_{-\infty}^{+\infty} (e^{ixh} - e^{-ixh})^{k+1}\, dF(x) = 2^{k+1}(-1)^{(k+1)/2}\int_{-\infty}^{+\infty} (\sin xh)^{k+1}\, dF(x)$$

$$= 2^{k+1}(-1)^{(k+1)/2}\int_0^{\infty} (\sin xh)^{k+1}\, dG(x) = o(h^k),$$

so that

$$\int_0^{1/h} (\sin xh)^{k+1}\, dG(x) = o(h^k)$$

$$(7) \qquad \int_0^{1/h} x^{k+1}\, dG(x) = o(h^{-1})$$

$$(8) \qquad \int_{1/2h}^{1/h} dG(x) = o(h^k).$$

On the other hand,

$$i^{-k}\frac{\Delta_h^k\varphi(0)}{(2h)^k} = \int_{-\infty}^{+\infty}\left(\frac{\sin xh}{xh}\right)^k x^k\,dF(x) = \int_0^{\infty}\left(\frac{\sin xh}{xh}\right)^k x^k\,dH(x)$$

$$= \int_0^{1/h} + \int_{1/h}^{\infty} = A_h + B_h,$$

say. Here

$$|\,B_h\,| \leq h^{-k}\int_{1/h}^{\infty} dG(x) = h^{-k}\left[\int_{1/h}^{2/h} + \int_{2/h}^{4/h} + \cdots\right]$$

$$= h^{-k}\left[o(h^k) + o\left(\frac{h}{2}\right)^k + \cdots\right] = o(1),$$

by (8). Since

$$\left(\frac{\sin u}{u}\right)^k = \{1 + O(u^2)\}^k = \{1 + O(u)\}^k = 1 + O(u)$$

for small $u$, we immediately obtain

$$A_h - \int_0^{1/h} x^k\,dH(x) = \int_0^{1/h} O(hx^{k+1})\,dG(x) = o(1),$$

by (7). Collecting the results, we see that

$$i^{-k}\frac{\Delta_h^k\varphi(0)}{(2h)^k} - \int_0^{1/h} x^k\,dH(x) = i^{-k}\frac{\Delta_h^k\varphi(0)}{(2h)^k} - \int_{-1/h}^{1/h} x^k\,dF(x) = o(1),$$

which completes the proof of Theorem 2.

One more remark. By Theorem 2, the existence of the first moment is equivalent to the existence of the first symmetric derivative

$$D_{(1)}\varphi(0) = \lim_{h\to 0}\,[\varphi(h) - \varphi(-h)]/2h.$$

In Theorem 1 we have a corresponding result for ordinary first derivative

$$\varphi'(0) = \lim_{h\to 0}\,[\varphi(h) - \varphi(0)]/h.$$

There is no discrepancy here since at every point where $\varphi$ is smooth the two notions of derivative are equivalent.

## REFERENCES

[1] A. Zygmund, *Trigonometrical series*, Warszawa-Lwów, 1935, p. 108.

[2] R. Fortet, "Calcul des moments d'une fonction de répartition à partir de sa caractéristique," *Bull. des Sci. Math.*, Vol. 68 (1944), pp. 117–131.

[3] Harald Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946, p. 90.

# A LOWER BOUND FOR THE VARIANCE OF SOME UNBIASED SEQUENTIAL ESTIMATES

By D. Blackwell and M. A. Girshick

*Howard University* and *Bureau of the Census*

Consider a sequence of independent chance variables $x_1, x_2, \cdots$ with identical distributions determined by an unknown parameter $\theta$. We assume that $E\,x_i = \theta$ and that $W_k = x_1 + \cdots + x_k$ is a sufficient statistic for estimating $\theta$ from $x_1, \cdots, x_k$. A sequential sampling procedure is defined by a sequence of mutually exclusive events $S_k$ such that $S_k$ depends only on $x_1, \cdots, x_k$ and $\Sigma P(S_k) = 1$. Define $W = W_k$ and $n = k$ when $S_k$ occurs. In a previous paper by one of the authors [1] it was shown that if $S_k = W_k C(S_1 + \cdots + S_{k-1})$, (where $C(A)$ denotes the event that $A$ does not occur), the function $V(W, n) = E(x_1 \mid W, n)$ is an unbiased estimate of $\theta$, and $\sigma^2(V) \leq \sigma^2(x_1)$. It is the purpose of this note to obtain a lower bound for $\sigma^2(V)$. Our result is:

Theorem i. $\sigma^2(V) \geq \dfrac{\sigma^2(x_1)}{E(n)}$.

We remark that the lower bound is actually attained in the classical case of samples of constant size $N$. For in this case, (see [1]), $V = E(x_1 \mid W_N) = W_N/N$. In fact we shall show that in a sense this is the only case in which the lower bound is attained.

The proof of Theorem I depends on certain properties of sums of independent chance variables. These, formulated more generally than is required for the proof of Theorem I, are given in

Theorem ii. *Let $x_1, x_2, \cdots$ be independent chance variables with identical distributions, having mean $\theta$ and variance $\sigma^2(x_1)$. Let furthermore $\{S_k\}$ be any sequential test for which $E(n)$ is finite. Let $W = x_1 + \cdots + x_k$ when $n = k$. Then*

(a) $\sigma^2(W - \theta n) \leq \sigma^2(x_1)\, E(n)$.

(b) *If $\sigma^2(n)$ is finite, the equality sign holds in (a).*

(c) $E[x_1(W - \theta n)] = \sigma^2(x_1)$.

Proof of (a). Write $y_i = x_i - \theta$, and define $Y = y_1 + \cdots + y_k$ when $n = k$. By definition,

$$\sigma^2(W - \theta n) = \sum_{k=1}^{\infty} \int_{S_k} (y_1 + \cdots + y_k)^2\, dP. \tag{1}$$

To prove (a), we must verify that the series on the right of expression (1) converges and has sum $\leq \sigma^2(x_1)E(n)$. Now

$$\begin{aligned} &\sum_{k=1}^{N} \int_{S_k} (y_1 + \cdots + y_k)^2\, dP \\ &\leq \sum_{k=1}^{N-1} \int_{S_k} (y_1 + \cdots + y_k)^2\, dP + \int_{n \geq N} (y_1 + \cdots + y_N)^2\, dP \\ &= \sum_{k=1}^{N} \int_{n \geq k} y_k^2\, dP + 2 \sum_{k=2}^{N} \int_{n \geq k} y_k(y_1 + \cdots + y_{k-1})\, dP. \end{aligned} \tag{2}$$

Since the event $\{n \geq k\}$ is independent of $y_k$, each term in the second sum vanishes and the first sum becomes

$$\begin{aligned}\sum_{k=1}^{N} \int_{\{n\geq k\}} y_k^2\, dP &= \sigma^2(x_1) \sum_{k=1}^{N} P\{n\geq k\} \\ &= \sigma^2(x_1)[P\{n = 1\} + 2P\{n = 2\} + \cdots NP\{n = N\} \\ &\qquad + NP\{n > N\}] \leq \sigma^2(x_1)E(n).\end{aligned} \tag{3}$$

This establishes Theorem II(a).

Proof of Theorem II(b). Write $z_i = | y_i |$ and let $Z = z_1 + \cdots + z_k$ when $n = k$. From (a) it follows that $\sigma^2[(Z - nE(z_i)]$ is finite. If in addition, $\sigma^2(n) < \infty$ then $E(Z^2) < \infty$. Thus the series

$$\sum_{k=1}^{\infty} \int_{S_k} (z_1 + \cdots + z_k)^2\, dP = \sum_{1\leq i,j\leq k<\infty} \int_{S_k} z_i z_j\, dP \tag{4}$$

converges, so that the series

$$\sum_{1\leq i,j,\leq k<\infty} \int_{S_k} y_i y_j\, dP \tag{5}$$

converges absolutely. The terms of the latter series may be arranged to yield

$$\text{(A)}: \sum_{k=1}^{\infty} \int_{S_k} (y_1 + \cdots + y_k)^2\, dP = \sigma^2\,(W - \theta n)$$

or to yield

$$\text{B}: \sum_{k=1}^{\infty} \int_{\{n\geq k\}} y_k^2\, dP + 2\sum_{k=2}^{\infty} \int_{\{n\geq k\}} y_k(y_1 + \cdots + y_{k-1})\, dP = \sigma^2(x_1)E(n).$$

This proves Theorem II(b).

Proof of Theorem II(c). It follows from Theorem II(a) that $Ex_1(W - \theta n)$ is finite. If we show that

(6) $E(W - \theta n \mid x_1) = x_1 - \theta$, i.e. $E(Y \mid y_1) = y_1$, it will follow [1] that

(7) $E[x_1(W - \theta n)] = E[x_1(x_1 - \theta)] = \sigma^2(x_1)$.

To verify (6), it is sufficient to show that if $f(x_1)$ is the characteristic function of an event depending only on $x_1$ (i.e. $f(x_1) = 1$ when the event occurs, $f(x_1) = 0$ otherwise)

$$E(fy_1) = E(fY). \tag{8}$$

Write $\phi_1 = 0$, $\phi_i = f\cdot(y_2 + \cdots + y_i)$, $i \geq 2$.
Then it easily verified that

$$E(\phi_j \mid x_1, \cdots, x_i) = \phi_i \text{ for } j \geq i \tag{9}$$

$$E\phi_i \leq \sum_{k=1}^{i} | y_k | \tag{10}$$

$$E(\phi_i) = 0. \tag{11}$$

Hence it follows [2] that $E\phi = 0$ where $\phi = \phi_i$ when $n = i$. In our case $\phi = fY - fy_1$, and $E\phi = 0$ yields (6). This completes the proof of Theorem II.

PROOF OF THEOREM I. In [1] it is proved that $E(x_1(W - \theta n)) = E[V(W - \theta n)]$. Hence employing Theorem II we get

$$\sigma^2(x_1) = E[V(W - \theta n)] = \sigma(V)\sigma(W - \theta n)\rho \tag{12}$$

where $\rho$, $(0 \leq \rho \leq 1)$, is the coefficient of correlation between $V$ and $W - \theta n$. Substituting for $\sigma(W - \theta n)$ we get

$$\begin{aligned} \sigma^2(x_1) &\leq \sigma(V)\sigma(x_1)\sqrt{E(n)\,\rho} \\ &\leq \sigma(V)\sigma(x_1)\sqrt{E(n)}. \end{aligned} \tag{13}$$

Solving for $\sigma(V)$ we finally obtain

$$\sigma^2(V) \geq \frac{\sigma^2(x_1)}{E(n)} \tag{14}$$

which proves Theorem I.[1]

If $\sigma^2(n)$ is finite, the equality sign in (14) will hold if and only if $\rho = 1$. We shall now prove the following.

THEOREM III. *Let $N$ be the minimum value of $n$ for which $P(n = N) \neq 0$. Then, a necessary and sufficient condition that $\rho = 1$ is that $P(n = N) = 1$.*

PROOF. The sufficiency of this condition follows from the fact that if $P(n = N) = 1$, $V = W/N$. To prove the necessity of this condition, we observe that if $\rho = 1$, $V$ is a linear function of $W - n\theta$. That is,

$$V = \alpha(W - n\theta) + \beta. \tag{15}$$

Now, since $EV = \theta$ and $E(W - n\theta) = 0$, it follows that $\beta = \theta$. Also, since by hypothesis $\sigma^2(V) = \sigma^2(x_1)/E(n)$ and $\sigma^2(W - n\theta) = \sigma^2(x_1)E(n)$, it follows that $\alpha = 1/\mathrm{E}(n)$. Hence the estimate $V$ is given by

$$V = \frac{W - n\theta}{E(n)} + \theta. \tag{16}$$

[1] Under certain regularity conditions Cramér has obtained the inequality

$$\sigma^2(x) \geq 1/E\left(\frac{\partial \log f}{\partial \theta}\right)^2$$

where $\mathrm{f} = \mathrm{f}(x, \theta)$ is the density function of $x$ ([3], p. 475). Thus with the same regularity conditions, our inequality yields

$$\sigma^2(V) \geq 1/E(n)E\left(\frac{\partial \log f}{\partial \theta}\right)^2,$$

which is a special case of the results presented by J. Wolfowitz in this issue of the *Annals*.

Let $N$ be defined as above. We note that $N < \infty$ since by hypothesis $E(n) < \infty$. Let $V_N$ be the estimate of $\theta$ when the sequential test terminates with $n = N$. Then $V_N = W/N$. Substituting this value in (16) we get

$$\frac{W}{N} - \theta = \frac{N}{E(n)}\left[\frac{W}{N} - \theta\right]. \tag{17}$$

We exclude the trivial case where $W \equiv N\theta$. Then (16) yields $E(n) = N$. That is $P(n = N) = 1$. This proves the theorem.

We remark that $N$ may be a function of $\theta$ but for a fixed $\theta$, $n = N$ is fixed when $\rho = 1$.

## REFERENCES

[1] D. Blackwell, "Conditional expectation and unbiased sequential estimation." Submitted to *Annals of Math. Stat.*

[2] D. Blackwell and M. A. Girshick, "On sums of sequences of independent chance vectors, with applications to the random walk in k dimensions," *Annals of Math. Stat.*, Vol. 17 (1946).

[3] Harald Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946.

# AN EXTENSION TO TWO POPULATIONS OF AN ANALOGUE OF STUDENT'S $t$-TEST USING THE SAMPLE RANGE

By John E. Walsh

*Princeton University*

**1. Summary.** The modified $t$-test considered by Daly[1] (see [1]) is used to develop one-sided significance tests to decide whether the mean of a new normal population exceeds the mean of an old normal population having the same variance. Significance tests are also developed to decide whether the mean of the new population is less than the mean of the old population. These tests require very little computation for their application and are approximately as powerful as the most powerful tests of these hypotheses.

**2. Introduction.** Let $r_1, \cdots, r_n$, $(n \leq 10)$, be independently distributed according to a normal distribution with zero mean and unit variance. Let $r_{(u)}$ denote the $u$th largest of the $r$'s. Then Daly has shown how to determine numbers $g_\alpha$ such that

$$\begin{aligned} &Pr[\bar{r}/(r_{(n)} - r_{(1)}) > g_\alpha] = \alpha \\ &Pr[\bar{r}/(r_{(n)} - r_{(1)}) < -g_\alpha] = \alpha. \end{aligned} \tag{1}$$

This note will use these relations to develop easily applied significance tests to decide whether the mean $\nu$ of a new normal population exceeds the mean $\mu$ of

[1] This problem is also considered by Lord in [2]. This note was in proof when [2] appeared.

an old normal population with the same variance. Significance tests are also developed to test $\nu < \mu$. The simplest case considered is that of testing a new sample value $x$ on the basis of $n$ past sample values $y_1, \cdots, y_n$. Then the significance test at significance level $\alpha$ to decide whether $\nu$ exceeds $\mu$ consists in accepting $\nu > \mu$ if

$$x > \bar{y} + g_\alpha \sqrt{n+1}[y_{(n)} - y_{(1)}],$$

where $y_{(u)}$ is the $u$th largest of $y_1, \cdots, y_n$.

The significance test of $\nu < \mu$ consists in accepting $\nu < \mu$ if

$$x < \bar{y} - g_\alpha \sqrt{n+1}\, [y_{(n)} - y_{(1)}].$$

These tests are generalized to the case in which $x$ is the mean of a sample of size $r$ from the new population, each of $y_1, \cdots, y_n$ is the mean of a sample of size $s$ from the old population, and $z$ is the mean of a sample of size $t$ from the old population. Then the tests at significance level $\alpha$ take the form

$$\begin{aligned} &\textit{Accept } \nu > \mu \textit{ if } x > (1 - C_1)\bar{y} + C_1 z + g_\alpha[y_{(n)} - y_{(1)}]; \\ &\textit{Accept } \nu < \mu \textit{ if } x < (1 - C_1)\bar{y} + C_1 z - g_\alpha[y_{(n)} - y_{(1)}], \end{aligned} \tag{2}$$

where $C_1$ is a given constant which is selected by the person applying the test. The introduction of the terms $z$ and $C_1$ allows less reliable past information to be utilized by lumping it together in the $z$ term and using the constant $C_1$ to weight this information according to its relative importance with respect to the $y$'s.

The power of test (2) is compared with that of the corresponding Student $t$-test for the case $C_1 = 0$ and $n \leq 10$. In this comparison the quantities $x, y_1, \cdots, y_n$ are considered to be the given sample values which are used for the test, that is, the quantities from which the means $x, y_1, \cdots, y_n$ were formed are not given. It is found that the power of the Student $t$-test is only slightly greater than that of the corresponding test (2). For the cases considered, however, it is well known that the most powerful test of $\nu > \mu$ using the quantities $x, y_1, \cdots, y_n$ is the appropriate Student $t$-test. Similarly for testing $\nu < \mu$. Thus the tests (2) considered are approximately as powerful as the most powerful tests of $\nu > \mu$ and $\nu < \mu$ which use $x, y_1, \cdots, y_n$.

Examination of (2) shows that the amount of computation required for the application of one of these tests is small. Consequently the tests (2) have the desirable properties of being easily computed and nearly as powerful as any tests which could be used for the given hypotheses. This suggests their use in repetitive testing procedures which are concerned with the testing of the mean of a new sample on the basis of the means of previous samples.

**3. Statement of tests.** In this section three significance tests of increasing generality are stated. It is to be observed that each test is a particular example of the test following it so that tests ($A$) and ($B$) are special cases of test ($C$).

The reason for stating tests $(A)$ and $(B)$ is that these tests have a much simpler appearance and will cover most cases of practical application.

$(A)$. Let each of $x, y_1, \cdots, y_n$ represent the mean of a sample of size $r$; let the values of the sample whose mean is $x$ have the distribution $N(\nu, \sigma^2)$ and the values of the samples whose means are $y_1, \cdots, y_n$ have distribution $N(\mu, \sigma^2)$, where the notation $N(\xi, \sigma^2)$ denotes the normal distribution with mean $\xi$ and variance $\sigma^2$. Then the significance test of $\nu > \mu$ at significance level $\alpha$ is

$$\textit{Accept} \quad \nu > \mu \quad \textit{if} \quad x > \bar{y} + g_\alpha \sqrt{\frac{n+1}{r}}\,[y_{(n)} - y_{(1)}].$$

The significance test to decide whether $\nu < \mu$ is

$$\textit{Accept} \quad \nu < \mu \quad \textit{if} \quad x < \bar{y} - g_{(\alpha)} \sqrt{\frac{n+1}{r}}\,[y_{(n)} - y_{(1)}].$$

$(B)$. Let $x$ equal the mean of $r$ sample values from $N(\nu, \sigma^2)$ and each of $y_1, \cdots, y_n$ equal the mean of $s$ sample values from $N(\mu, \sigma^2)$. The significance test for $\nu > \mu$ at significance level $\alpha$ is

$$\textit{Accept} \quad \nu > \mu \quad \textit{if} \quad x > \bar{y} + g_\alpha \sqrt{\frac{n}{r} + \frac{1}{s}}\,[y_{(n)} - y_{(1)}].$$

The test of $\nu < \mu$ is given by

$$\textit{Accept} \quad \nu < \mu \quad \textit{if} \quad x < \bar{y} - g_\alpha \sqrt{\frac{n}{r} + \frac{1}{s}}\,[y_{(n)} - y_{(1)}].$$

$(C)$. Let $x$ equal the mean of $r$ sample values from $N(\nu, \sigma^2)$, each of $y_1, \cdots, y_n$ equal the mean of a sample of size $s$ from $N(\mu, \sigma^2)$, $z$ equal the mean of a sample of size $t$ from $N(\mu, \sigma^2)$, and $C_1$ be a given constant value. Then the significance test of $\nu > \mu$ at significance level $\alpha$ is

*Accept* $\nu > \mu$ *if*

$$x > (1 - C_1)\bar{y} + C_1 z + [y_{(n)} - y_{(1)}]g_\alpha \cdot \sqrt{\left(\frac{1}{r} + \frac{C_1^2}{t}\right)\left(n + \frac{(1 - C_1)^2}{s\left(\frac{1}{r} + \frac{1}{t}C_1^2\right)}\right)}.$$

The significance test to decide whether $\nu < \mu$ is

*Accept* $\nu < \mu$ *if*

$$x < (1 - C_1)\bar{y} + C_1 z - [y_{(n)} - y_{(1)}]g_\alpha \cdot \sqrt{\left(\frac{1}{r} + \frac{C_1^2}{t}\right)\left(n + \frac{(1 - C_1)^2}{s\left(\frac{1}{r} + \frac{1}{t}C_1^2\right)}\right)}.$$

Values of $g_\alpha$ for $\alpha = .05$ are given in Table I. These values were listed by Daly in [1].[2]

[2] Values of $g_\alpha$ for $\alpha = .05, .025, .01, .005, .001.$ and $.0005$ are listed in Table 9 of [2] for sample sizes from 2 to 20.

**4. Derivation of tests.** As tests ($A$) and ($B$) are particular cases of test ($C$), it is sufficient to derive test ($C$).

TABLE I

*Estimated Values of $g_{.05}$*

| $n$ | $g_{.05}$ |
|---|---|
| 3 | .882 |
| 4 | .526 |
| 5 | .385 |
| 6 | .309 |
| 7 | .260 |
| 8 | .227 |
| 9 | .202 |
| 10 | .183 |

Let the quantities $x', y'_1, \cdots, y'_n, z'$ be defined by

$$x' = \frac{(x - \nu)\sqrt{r}}{\sigma}, \qquad y'_i = \frac{(y_i - \mu)\sqrt{s}}{\sigma}, \qquad (i = 1, \cdots, n),$$

$$z' = \frac{(z - \mu)\sqrt{t}}{\sigma}.$$

Then $x', y'_1, \cdots, y'_n, z'$ are independently distributed according to $N(0, 1)$. Define

$$r_u = \frac{1}{K_1}\left(K_1 y'_u - \sum_1^n y'_i + K_2 x' + K_2 C z'\right), \qquad (u = 1, \cdots, n).$$

It is easily verified that

$$E(r_u) = 0, \qquad E(r_u^2) = \frac{1}{K_1^2}[K_1^2 + (1 + C^2)K_2^2 - 2K_1 + n]$$

$$E(r_u r_v) = \frac{1}{K_1^2}[(1 + C^2)K_2^2 - 2K_1 + n], \qquad (u \neq v).$$

Thus, if $K_1$ and $K_2$ satisfy the equations

$$(3) \qquad \left(\sqrt{\frac{r}{s}} + C\sqrt{\frac{t}{s}}\right)K_2 + K_1 - n = 0$$

$$(1 + C^2)K_2^2 - 2K_1 + n = 0,$$

the $r_u$ will be independent of $\mu$ when $\mu = \nu$. Also they will be independently distributed according to $N(0, 1)$.

Rewriting the $r_u$ in terms of $x, y_1, \cdots, y_n, z$ one obtains

$$(4)\quad r_u = \frac{\sqrt{s}}{K_1\sigma}\left[K_1 y_u - \sum_1^n y_i + K_2\sqrt{\frac{r}{s}}\,x + K_2 C\sqrt{\frac{t}{s}}\,z + K_2\sqrt{\frac{r}{s}}\,(\mu - \nu)\right].$$

Using (3) the mean of the $r_u$ is found to be

$$\bar{r} = \frac{K_2\sqrt{r}}{K_1\sigma}\left[x - \left(1 + C\sqrt{\frac{t}{r}}\right)\bar{y} + C\sqrt{\frac{t}{r}}\,z - (\nu - \mu)\right].$$

Let $r_{(u)}$ denote the $u$th largest of $r_1, \cdots, r_n$. Then from (1)

$$\alpha = Pr[\bar{r}/(r_{(n)} - r_{(1)}) > g_\alpha] = Pr\left[\frac{K_2\sqrt{r}}{K_1}\left(x - \left(1 + C\sqrt{\frac{t}{r}}\right)\bar{y} + C\sqrt{\frac{t}{r}}\,z - (\nu - \mu)\right)\Big/(y_{(n)} - y_{(1)}) > g_\alpha\right]$$

It is easily proved from (3) that

$$\frac{K_1}{K_2\sqrt{r}} = \pm\sqrt{\frac{1 + C^2}{r}\left(n + \frac{(\sqrt{r} + C\sqrt{t})^2}{s(1 + C^2)}\right)}.$$

Choosing the positive sign, putting $C = -\sqrt{\frac{r}{t}}\,C_1$, and letting $\mu = \nu$ one obtains

$$Pr\left[x > (1 - C_1)\bar{y} + C_1 z + [y_{(n)} - y_{(1)}]g_\alpha \cdot \sqrt{\left(\frac{1}{r} + \frac{C_1^2}{t}\right)\left(n + \frac{(1 - C_1)^2}{s\left(\frac{1}{r} + \frac{1}{t}C_1^2\right)}\right)}\right] = \alpha,$$

verifying the first part of test ($C$). The second part of test ($C$) is verified by choosing the negative sign for $\frac{K_1}{K_2\sqrt{r}}$ (or by repeating the above argument using the second part of (1)).

**5. Power comparison with t-test.** Let $x, y_1, \cdots, y_n$ satisfy the conditions of test ($B$) in section 3. Then Student's $t$ using $x, y_1, \cdots, y_n$ is given by

$$t = \frac{[x - \bar{y} - (\nu - \mu)]}{\sqrt{\sum_1^n (y_i - \bar{y})^2}} \cdot \sqrt{\frac{n - 1}{s\left(\frac{1}{r} + \frac{1}{ns}\right)}}.$$

The Student $t$-test based on this value of $t$ furnishes the most powerful test of $\nu > \mu$ (and $\nu < \mu$) using $x, y_1, \cdots, y_n$. The purpose of this section is to show that test ($B$) has approximately the same power as this Student $t$-test for $n \leq 10$.

Daly has shown (see [1]) that if $r_1, \cdots, r_n$ are independently distributed according to $N(\xi, \sigma^2)$, then the test based on

$$(\bar{r} - \xi)/(r_{(n)} - r_{(1)})$$

has approximately the same power for testing $\xi > 0$ (and $\xi < 0$) as the corresponding Student $t$-test based on

$$t = \frac{(\bar{r} - \xi)\sqrt{n(n-1)}}{\sqrt{\sum_1^n (r_i - \bar{r})^2}} \tag{5}$$

for $n \leq 10$.

Using the notation of section 4 let

$$r_u = \frac{\sqrt{s}}{K_1}\left[K_1 y_u - \sum_1^n y_i + K_2\sqrt{\frac{r}{s}}\,x\right], \qquad (u = 1, \cdots, n),$$

where $\frac{K_1}{K_2} > 0$. Then from consideration of (4) with $C = 0$ it is seen that the $r_u$ are independently distributed according to $N(\xi, \sigma^2)$, where $\xi$ equals a positive constant times $(\nu - \mu)$. Following the derivations in section 4 with $C = 0$, it is seen that the test of $\xi > 0$ with this particular choice of the $r_u$ is identical with the test of $\nu > \mu$ given in $(B)$ of section 3. Similarly the test of $\xi < 0$ is identical with the test $(B)$ of $\nu < \mu$. Thus the test $(B)$ has approximately the same power for testing $\nu > \mu$ (and $\nu < \mu$) as the Student $t$-test based on the value of $t$ given in (5) if $n \geq 10$. Replacing the $r_u$ in (5) by their values in terms of $x, y_1, \cdots, y_n, n, r,$ and $s$, it is found that (5) becomes

$$t = \frac{[x - \bar{y} - (\nu - \mu)]}{\sqrt{\sum_1^n (y_i - \bar{y})^2}} \cdot \sqrt{\frac{n-1}{s\left(\frac{1}{r} + \frac{1}{ns}\right)}}.$$

This proves that test $(B)$ is approximately as powerful for testing $\nu > \mu$ and $\nu < \mu$ as the most powerful test based on the quantities $x, y_1, \cdots, y_n$ if $n \leq 10$. As test $(A)$ is a particular case of test $(B)$, these results also apply to test $(A)$.

## REFERENCES

[1] J. F. Daly, "On the use of the sample range in an a analogue of Student's $t$-test". *Annals of Math. Stat.*, Vol. 17 (1946), pp. 71-74.

[2] E. Lord, "The use of range in place of standard deviation in $t$-test," *Biometrika*, Vol. 34 (1947), pp. 41–67.

# ON THE NORM OF A MATRIX

By Albert H. Bowker
*University of North Carolina*

In studying the convergence of iterative procedures in matrix computation and in setting limits of error after a finite number of steps, Hotelling [1] used the square root of the sum of squares of the elements of a matrix as its norm. A wide class of functions exists which may be employed as norms in matrix calculation and substituted directly in the expressions derived by Hotelling. The purpose of this note is to make a few general remarks about this class of functions and to propose a new norm which appears to have some value in computation.

A function $\phi(A)$ of the elements of a real matrix $A$ may be termed a legitimate norm if it has the following four properties:

$$\phi(cA) = | c | \phi(A), \; c \text{ a scalar;} \tag{1}$$

$$\phi(A + B) \leqq \phi(A) + \phi(B), \text{ if } A + B \text{ is defined;} \tag{2}$$

$$\phi(AB) \leqq \phi(A)\phi(B), \text{ if } AB \text{ is defined;} \tag{3}$$

$$\phi(e_{ij}) = 1, \text{ where } e_{ij} \text{ is a fundamental unit matrix} \tag{4}$$

whose elements are all zero except the one in the $i$th row and $j$th column, whose value is unity. These four conditions are identical with the first four axioms of Rella [2], who has shown them to be independent. Properties (1), (2), and (3) are used directly in investigations of convergence and error, but the importance of property (4) is indicated by some of its immediate consequences. Clearly $e_i'Ae_j = a_{ij}$, where $e_i$ is a fundamental unit vector. From (3) and (4) it follows that $| a_{ij} | \leqq \phi(A)$ for all $i$ and $j$ and we have that

$$\max_{(ij)} | a_{ij} | \leqq \phi(A). \tag{5}$$

Thus $\phi(A)$ has the useful property that the norm of a matrix of errors exceeds or equals the maximum possible error. Since $\phi(A^m) \leqq \phi^m(A)$, it follows from (5) that the elements of $A^m$ will tend to zero as $m$ increases if $\phi(A) < 1$, a result which is useful in establishing convergence. Also $\phi(A) \geqq 0$.

One further consequence of (1) to (4) is of interest. Suppose $A$ is a square matrix and let $\lambda$ be any of its roots. Then there exists a non-null vector $x$ such that $Ax = \lambda x$. Now $\phi(\lambda x) = \lambda\phi(x) \leqq \phi(A)\phi(x)$ and we have

$$\lambda \leqq \phi(A). \tag{6}$$

Thus, every legitimate norm is an upper bound to the characteristic roots.

Clearly many functions exist which satisfy (1) to (4). The norm used by Hotelling is $N(A) = \sqrt{\sum_{i,j} a_{ij}^2}$. A new norm which may have some value is obtained as follows:

$$R(A) = \max_{(i)} R_i(A) \tag{7}$$

where

$$R_i(A) = \sum_j | a_{ij} | .$$

Clearly $R(cA) = | c | R(A)$. To show that $R$ satisfies (2), consider

$$R_i(A + B) = \sum_j | a_{ij} + b_{ij} | \leqq \sum_j | a_{ij} | + \sum_j | b_{ij} | \leqq R(A) + R(B).$$

Since the above inequality holds for all $i$,

$$R(A + B) \leqq R(A) + R(B).$$

Now $AB = || \sum_\alpha a_{i\alpha} b_{\alpha j} ||$

and

$$R_i(AB) = \sum_j | \sum_\alpha a_{i\alpha} b_{\alpha j} | \leqq \sum_j \sum_\alpha | a_{i\alpha} | \cdot | b_{\alpha j} |$$
$$\leqq \sum_\alpha | a_{i\alpha} | R_\alpha(B) \leqq R(B)R(A).$$

Hence $R(AB) \leqq R(A)R(B)$. Clearly $R(e_{ij}) = 1$. Similarly it may be shown that $C(A) = \max_{(j)} \sum_i | a_{ij} |$ also satisfies the conditions of a norm.

Since the convergence of an iterative procedure is often proved by the norm being less than one, since the norm appears in the upper bound for the error after a finite number of iterations, and since the norm of a matrix of errors is taken to indicate the magnitude of the errors, a reasonable method of choosing among several available legitimate norms is to select the smallest. It is natural to inquire whether an optimum norm in this sense exists; that is, is there a function $\phi^*(A)$ such that $\phi^*(A)$ possesses properties (1) through (4) and such that $\phi^*(A) \leqq \phi(A)$ for all other $\phi(A)$ satisfying these conditions. Assume such a $\phi^*(A)$ does exist. Clearly $\phi^*(A) = \phi^*(A')$, as, if either exceeded the other, the smaller could be taken as $\phi^*(A)$. Let $\Lambda^2$ be the largest root of $AA'$. Then by (6)

$$\Lambda^2 \leqq \phi^*(AA') \leqq \phi^{*2}(A) \text{ and } \Lambda \leqq \phi^*(A).$$

But Rella [2] has shown that $\Lambda$ possesses (1) to (4). Thus

$$\phi^*(A) = \Lambda.$$

But, for a row vector, $C(A) \leqq \Lambda$. Consequently, no minimal norm exists. It is interesting to note that a worst norm does exist, namely $P(A) = \sum_{i,j} | a_{ij} |$. Since $A = \sum_{i,j} e_{ij} a_{ij}$, $\phi(A) \leqq P(A)$. Clearly $P(A)$ satisfies (1) to (4) and hence is the worst possible legitimate norm.

In practical computation, the choice so far is between $N(A)$ and $R(A)$ (or $C(A)$). No general inequalities exist and it would probably be advisable to compute both. $R(A)$ may be less than $N(A)$ and indicate convergence when $N(A)$ fails to do so. Often $R(A)$ may be computed visually and convergence proved without computing the sum of squares of the elements.

The functions $N(A)$ and $R(A)$ may also be useful in finding a simple first approximation to $A^{-1}$. A sufficient condition that Hotelling's iterative method for finding the inverse of a matrix $A$ will converge is that the roots of $D = 1 - AC_0$ be less than one in absolute value where $C_0$ is a first approximation to $A^{-1}$. If the iterative procedure is to be carried out by a fully automatic computing machine such as the one described by Alt [3] it may be advisable to start with a rather poor first approximation which is easy to construct. If $A$ has positive roots and if $M$ is any upper bound to these roots and if $C_0$ is a matrix with diagonal elements equal to $1/M$ and zeros elsewhere, the iterative procedure will converge but the norm of $D$ will not necessarily be less than one. From (6), any legitimate norm may be taken as $M$.

Finally, it is interesting to point out the relation of this note to some work on the problem of finding upper bounds to the roots. In fact, the inequalities $\lambda \leqq N(A)$ and $\lambda \leqq R(A)$, which are consequences of (6), are Theorem 2 of Farnell [4] and Theorem 3 of Barankin [5] respectively.

## REFERENCES

[1] HAROLD HOTELLING, "Some new methods in matrix calculation," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 1-34.

[2] T. RELLA, "Über den absoluten Betrag von Matrizen," International Congress of Mathematicians at Oslo (1936).

[3] FRANZ L. ALT, "Multiplication of matrices," *Math. Tables and Aids to Compution*, Vol. 2 (1946), pp. 12-13.

[4] A. B. FARNELL, "Limits for the characteristic roots of a matrix," *Bull. Amer. Math. Soc.*, Vol. 50 (1944), pp. 789-794.

[5] EDWARD W. BARANKIN, "Bounds for the characteristic roots of a matrix," *Bull. Amer. Math. Soc.*, Vol. 51 (1945), pp. 767-770.

# DEFINITION OF THE PROBABLE DEVIATION

BY M. FRÉCHET

*Faculty of Science, University of Paris*

The probable deviation has recently been defined by E. J. Gumbel [1], [2] as the smallest of the intervals corresponding to the probability $\frac{1}{2}$. It so happened that the author was led to an equivalent definition starting from a general idea which may be applied to absolutely general cases and which, for this reason, might be of interest.

In recent years, the author has been occupied with a study of random elements of any nature (curves, surfaces, functions, qualitative elements), a study whose future seems promising, [3]. I gave a definition of the mean of such an element expressed by an abstract integral which, however, is only defined if the random element is situated in a metric vectorial (Wiener-Banach) space.[1] But[2] a still more general definition is valid if the random element is placed in any metric space. It consists of taking, as mean position of the random element $X$, a fixed (non-statistical) element $b = \bar{X}$ such that the function of $a$ which represents the mean $M(X, a)^2$ of the squared distance of $X$ to the fixed element $a$, is minimum for $a = b$. (In the case where $X$ and $a$ are numbers, and where $M(X)^2$ is finite, we know that this minimum is reached and that there is one, and only one, determination $b$ of $a$). This definition has the advantage of also defining the equiprobable position of $X$. This is a fixed element $c = \bar{\bar{X}}$ such that $M(X, a)$ is minimum for $c = a$. (If $X$ and $a$ are numbers, we know that this minimum is still reached, but may be so reached by several values of $\bar{\bar{X}}$).

Since reading Gumbel's paper, a still more general definition suggested itself.

[1] For the definition of metric vectorial spaces see [4].

[2] See Note 2, p. 503 of [4].

The expressions $M(X, a)$ and $\sqrt{M(X, a)^2}$ themselves may be considered as distances, but as distances of two random elements *taken together*. To each of these distances corresponds as minimum, when $a$ varies, a different "typical" function $\bar{\bar{X}}$ or $\bar{X}$ $\cdots$. Thus, without supposing anything about the space into which the different trials place $X$, we assume that we have defined a "deviation" of two random elements $X$, $Y$ taken together. We represent this function of two random variables by $(([X], [Y]))$, a notation which differs from the representation of the distance $(X, Y)$ of the two positions $X$ and $Y$ with respect to a single trial. The lower boundary of the deviation $(([X], [a]))$, a function of $a$, which is reached for $a = \widehat{X}$ defines a "typical" position $\widehat{X}$. Moreover, the value of this $(([X], [\widehat{X}]))$ may be considered as a measure or, at least, as a numerical ranging point of the dispersion of $X$.

Let us abandon these generalities. They hold especially if the element $X$ is a real valued random variable. Among the possible and reasonable[3] expressions for the deviation $(([X], [a]))$ of the numerical variate $X$ from a fixed number $a$, we may use the equiprobable value of $| X - a |$ which may be called the equiprobable deviation of $X$ from $a$. Thus we have, on one side, a new "typical value" of $X$ which will be a value of $a$ such that the equiprobable deviation of $X$ from $a$ is minimum, and a new measure of dispersion which is the value of this minimum and which might be called simply the equiprobable deviation of $X$.

In the case where $X$ has everywhere a continuous and finite density of probability $w(X)$ we find, as typical value, what Gumbel calls the "midvalue" and represents by $\xi$, and, as equiprobable deviation, what Gumbel calls the "probable deviation" and represents by $\zeta$.

We may also consider the discontinuous case, which was given as a problem to candidates of the "Certificat d'Etudes Supérieures de Calcul des Probabilités, Option Statistique Mathématique, Session May-June, 1944." They had to solve various questions of which I cite the beginning below:

"Consider $n$ real numbers $x_1 \leqq x_2 \leqq \cdots \leqq x_n$ and represent, by $E_a$, a median value of the deviations $| x_k - a |$ of the numbers $x_k$ and $a$. If $a$ varies, $E_a$ has a minimum $E$ which is reached by one or several values $A$ of $a$.

1) Explain, in a few words, the meaning of the values $E$ and $A$.

2) For simplicity's sake, suppose that $n$ is odd ($n = 2r + 1$). How should $E$ and $A$ be calculated practically? (To find the answer, investigate first how $E_a$ varies if $a$ varies only slightly).

3) In the case where $n = 4s + 3$ ($s$ is an integer equal to, or larger than, zero) show that $E \leqq \dfrac{(q_3 - q_1)}{2}$

where $$q_1 = x_{s+1}, \quad q_3 = x_{n-s}."$$

The study of this new typical value and of this new equiprobable deviation has the advantage that their determination is very rapid and requires hardly

[3] See the *Remark* at end of note.

any calculations. However, we have to note an important inferiority of the equiprobable deviation of $X$ compared to the mean and the standard deviations of $X$. If one or the other of the last two deviations is zero, $X$ is a fixed number (except for the case of the probability zero). This property seems requested by the intuitive meaning which we attribute to the dispersion, and to every measure or any mark of it. Now, the equiprobable deviation lacks this property. If, for instance, $X$ has only three values: 0, 2, 1, the first two with the probability 0.249, and the last with the probability 0.502, the equiprobable deviation of $X$ will be zero, whereas $X$ will be equal to its typical value 1 only with a probability of 0.502, and not with a probability equal to unity. The same holds for any distribution for which there is a point with probability exceeding $\frac{1}{2}$.

*Remark.* The definitions of the mean and of the equiprobable position become meaningless in the case that $M(X, a)$, or $M(X, a)^2$, is infinite. However, we succeeded in surmounting the difficulty, and to reach definitions which are valid even in this case. If $X$ is a number, the new definitions become equivalent to the classical definitions of the mean and equiprobable value. The proofs are given in two recent articles [5], [6].

## REFERENCES

[1] E. J. GUMBEL, "Definition of the probable error," *Annals of Math. Stat.*, Vol. 13 (1942)' p. 110-111.

[2] E. J. GUMBEL, "Probable deviation," *Stat. Jour.*, City College, (N. Y.), Vol. 6 (1943), pp. 25-26.

[3] M. FRÉCHET, "L'intégrale abstraite d'une fonction abstraite d'une variable abstraite et son application à la moyenne d'un élément aléatoire de nature quelconque," *Revue Scientifique*, Vol. 82 (1944), pp. 483-512.

[4] M. FRÉCHET, *Les Espaces Abstraites*, Gauthier-Villars, Paris, 1928, pp. 125-141.

[5] M. FRÉCHET, "Les éléments aléatoires de nature quelconque," *Ann. Inst. H. Poincaré*, 1947.

[6] M. FRÉCHET, "Nouvelles définitions de la valeur moyenne et des valeurs probables d'un nombre aléatoire," *Ann. Univ. de Lyon*, 1947.

# THE GENERAL RELATION BETWEEN THE MEAN AND THE MODE FOR A DISCONTINUOUS VARIATE

BY M. FRÉCHET

*Faculty of Science, University of Paris*

Dr. Gumbel has pointed out that one of the author's arguments employed in several particular cases (see [1]) can be employed in a general case which includes them and leads to the following result: If a statistical variate $R$ has only positive entire values differing from zero, and if its mean value $\bar{R}$ is smaller than, or equal to, unity, the same holds for its equiprobable value $\bar{\bar{R}}$ and its mode $\bar{R}$. There are two generalizations of this result which might be of interest:

1) On the one hand, the author has shown [2] that, if a variate $R$ can only have values (entire or not) equal to, or larger than, zero, its equiprobable value

$\bar{\bar{R}}$ is, at most, equal to twice its mean value $\bar{R}$, and the inequality $\bar{\bar{R}}/\bar{R} \leqq 2$ cannot be improved which means that the upper boundary of the first member is exactly equal to (and not less than) two. The equality is reached when $R$ has only two values of equal probability, one of them being zero.

2) On the other hand, if $R$ is an integer positive variate equal to, or larger than zero, it can be proven that, if $\bar{R} \leqq \alpha$, we have

$$\check{R} \leqq \frac{\alpha(\alpha + 3)}{2}\,. \tag{1}$$

Here, $\bar{R}$ and $\check{R}$ stand for the mean and for the mode of $R$ respectively, and $\alpha$ is a positive integer differing from zero. For example: if $R$ is the number of repetitions of an event with probability $p$, we have, for $n$ trials, $\bar{R} = np$, whence, if $\alpha$ is the first integer number equal to, or larger than, $\bar{R}$ we have the inequality (1) for the most probable number of repetitions. Naturally, this inequality only has an interest if the second member of (1) is smaller than $n$ which means that

$$\alpha(\alpha + 3) < 2n.$$

This presupposes

$$2n > np(np + 3)$$

or

$$n < \frac{2 - 3p}{p^2}$$

and, since $n$ must be positive,

$$p < \tfrac{2}{3}.$$

To prove the inequality (1), let us write $\omega_\nu$ for the probability that $R = \nu$. We have

$$\sum_0^\infty \omega_\nu = 1; \qquad \sum_0^\infty \nu\omega_\nu = \bar{R} \leqq \alpha$$

whence

$$\sum_0^{\alpha-1} (\alpha - \nu)\omega_\nu \geqq \sum_{\alpha+1}^\infty (\nu - \alpha)\omega_\nu. \tag{2}$$

Let the mode be

$$\check{R} = \beta$$

then

$$\omega_\beta \geqq \omega_\nu\,;\ \nu = 0, 1, 2, \cdots$$

and the first member in (2) is bounded by

$$\frac{\alpha(\alpha + 1)}{2}\,\omega_\beta \geqq \sum_0^{\alpha-1} (\alpha - \nu)\omega_\nu. \tag{3}$$

Now, either $\alpha < \beta$ or $\beta \leqq \alpha$. In the first case the second member in (2) leads to

$$\sum_{\alpha+1}^{\infty} (\nu - \alpha)\omega_\nu \geqq (\beta - \alpha)\omega_\beta \tag{4}$$

since the second member in (4) is one of the terms occurring in the sum. The same inequality holds in the second case, $\beta \leqq \alpha$, hence it holds generally. It follows from (2), (3), and (4) that

$$\frac{\alpha(\alpha + 1)}{2}\,\omega_\beta \geqq (\beta - \alpha)\omega_\beta .$$

The probability $\omega_\beta$ is certainly different from zero, since $\sum_0^\infty \omega_\nu = 1$. Consequently

$$\beta - \alpha \leqq \frac{\alpha(\alpha + 1)}{2}$$

or

$$\beta \leqq \frac{\alpha(\alpha + 3)}{2}$$

as stated in (1).

The equality in (1) is possible only if, from (3),

$$\alpha(\omega_\beta - \omega_0) + (\alpha - 1)(\omega_\beta - \omega_1) + \cdots + (\omega_\beta - \omega_{\alpha-1}) = 0$$

and from (4)

$$\omega_{\alpha+1} + 2\omega_{\alpha+2} + \cdots + (\beta - \alpha)\omega_\beta + \cdots = (\beta - \alpha)\omega_\beta$$

whence

$$\omega_0 = \omega_1 = \cdots = \omega_\beta = \cdots = \omega_{\alpha-1} \tag{5}$$

and

$$\omega_{\alpha+1} = \omega_{\alpha+2} = \cdots = 0. \tag{5'}$$

The existence of the exceptional case proves that the inequality (1) cannot be improved by replacing the second member by a smaller function of $\alpha$. In the exceptional case, the only possible values of $R$ are

$$R = 0, 1, 2, \cdots, \alpha - 1, \alpha, \beta,$$

and all values, except perhaps $\alpha$, are equiprobable. The probability $\omega_\alpha$ may be, but need not be, equal to $\omega_\beta$ .

Moreover

$$\beta = \frac{\alpha(\alpha + 3)}{2} \geqq \alpha \tag{6}$$

and $\beta = \alpha$ is possible only if $\alpha = \beta = 0$ whence, from (5), $\omega_\nu = 0$ except for $\nu = 0$ which means that $R$ only has one value equal to zero. Except for this trivial case, we have in the exceptional case $\beta > \alpha$, and there are $\alpha + 2$ possible values for $R$. Then we must have

$$\omega_\beta \geqq \omega_\alpha; \qquad \sum_0^\alpha \omega_\nu + \omega_\beta = 1$$

whence

$$(\alpha + 1)\omega_\beta + \omega_\alpha = 1$$

and, from (5),

$$\alpha \geqq \bar{R} = \omega_\beta \sum_1^{\alpha-1} \nu + \beta\omega_\beta + \alpha\omega_\alpha = \omega_\beta\left(\frac{\alpha(\alpha-1)}{2} + \frac{\alpha(\alpha+3)}{2}\right) + \alpha\omega_\alpha$$

$$= \alpha((\alpha + 1)\omega_\beta + \omega_\alpha)$$

whence

$$\bar{R} = \alpha. \tag{7}$$

From

$$1 = (\alpha + 1)\omega_\beta + \omega_\alpha \geqq (\alpha + 2)\omega_\alpha$$

follows

$$\omega_\alpha \leqq \frac{1}{\alpha + 2}; \qquad \omega_\beta = \frac{1 - \omega_\alpha}{\alpha + 1}. \tag{8}$$

These conditions (5), (5′), and (7) are necessary and sufficient for the existence of the exceptional case.

If the equality in (1) is excluded, the mode $\beta$ and the smallest integer number $\alpha$ which is equal to, or larger than, the mean, are related by

$$\beta \leqq \frac{\alpha(\alpha + 3)}{2} - 1 = \frac{\alpha^2 + 3\alpha + 2}{2}. \tag{9}$$

As shown before, this general inequality, valid for any discontinuous variate, which can assume only non-negative integer values, cannot be improved without assuming specific properties of the distribution.

## REFERENCES

[1] M. Fréchet, *Les probabilités associées à un système d'evénements compatibles et dependents*, Hermann et Cie., 1943, Part II, p. 5.

[2] M. Fréchet, "Comparaison de diverses mesures de la dispersion", *Rev. de l'Inst. International de Stat.*, Vol. 8 (1940), p. 5.

# NOTE ON DIFFERENTIATION UNDER THE EXPECTATION SIGN IN THE FUNDAMENTAL IDENTITY OF SEQUENTIAL ANALYSIS

By T. E. Harris

*Princeton University*

Let $z$ be any chance variable and $z_1, z_2, z_3, \cdots$ a sequence of independent chance variables, each with the same distribution as $z$. Let $Z_N = z_1 + z_2 + \cdots + z_N$. Let $\phi(t) = Ee^{zt}$ for all complex $t$ for which the latter exists. Let $S_1$, $S_2, \cdots$ be a sequence of mutually exclusive events such that $S_j$ depends only on $z_1, z_2, \cdots, z_j$, and $\sum_{j=1}^{\infty} P(S_j) = 1$. Let the chance variable $n$ be defined as $n = j$ when $S_j$ occurs. Blackwell and Girshick [1], generalizing a result of Wald [2], showed that if there is a positive constant $M$ such that

$$(1) \qquad | Z_N | < M \text{ when } n > N$$

then the identity

$$(2) \qquad E\{e^{Z_n t}(\phi(t))^{-n}\} = 1$$

holds for all complex $t$ for which $\phi(t)$ exists and $| \phi(t) | \geq 1$. Wald [3] established conditions, including the existence of $\phi(t)$ for all real $t$, under which (2) may be differentiated under the expectation sign an unlimited number of times.

Without assuming the existence of $\phi(t)$ for a real $t$-interval the following result holds: *If (1) is true and if $E(z^k)$ and $E(n^k)$ are both finite, $k$ a positive integer, then*

$$(3) \qquad E\left\{\frac{d^k}{ds^k}[e^{Z_n is}(\phi(is))^{-n}]_{s=0}\right\} = 0$$

*where $i = \sqrt{-1}$ and $s$ is real.* Certain identities, obtained by differentiating (2) and putting $t = 0$, can also be obtained from (3). For example, if $En = 0$, and if $En^2$ and $Ez^2$ both exist then $EZ_n^2 = Ez^2En$.

Let $P_N = P(n \leq N)$; $p_N = P(n = N)$. Let $H(j, Z_j)$ and $F(N, Z_N)$ be the conditional cumulatives of $Z_j$ and $Z_N$ for $n = j$ and $n > N$ respectively. Now (2) was derived by Wald [2], p. 285, from a relation, valid whenever $\phi(t)$ exists, which in the present notation becomes

$$(4) \qquad \sum_{j=1}^{N} p_j \int_{-\infty}^{\infty} (\phi(t))^{-j} e^{Z_j t}\, dH(j, Z_j) + \frac{(1 - P_N)}{(\text{¢}(t))^N} \int_{-\infty}^{\infty} e^{Z_N t}\, dF(N, Z_N) = 1.$$

Examination of Wald's derivation of (4) shows it to be valid under the present hypotheses. Now the finiteness of $E(z^k)$ clearly implies that of $E(Z_j^k | n = j)$. Also, since $F(N, Z_N)$ is constant outside the interval $[-M, M]$, the integral $\int_{-\infty}^{\infty} Z_N^k\, dF(N, Z_N)$ is finite. Hence we may set $t = is$ in (4) and differentiate

$k$ times, obtaining for all real $s$

$$\begin{aligned}(5)\qquad &\sum_{j=1}^{N} p_j \int_{-\infty}^{\infty} \frac{d^k}{ds^k}[(\phi(is))^{-j}\, e^{Z_j is}]\, dH(j, Z_j) \\ &+ (1 - P_N) \sum_{r=0}^{k} \binom{k}{r} \frac{d^r}{ds^r}\,[(\phi(is))^{-N}]\cdot \int_{-\infty}^{\infty} (iZ_N)^{k-r}\, e^{Z_N is}\, dF(N, Z_N) = 0.\end{aligned}$$

The derivatives of $(\phi(is))^{-N}$ are sums of terms of the form $Q(N)\cdot(\phi(is))^{-N-r}$ times terms independent of $N$, where $Q(N)$ is a polynomial in $N$ of degree $\leq k$. For any $r \leq k$,

$$\lim_{N\to\infty} |\,(1 - P_N)N^r\,| = \lim_{N\to\infty} \left| N^r \sum_{j=N+1}^{\infty} p_j \right| \leq \lim_{N\to\infty} \left| \sum_{j=N+1}^{\infty} j^k p_j \right| = 0,$$

since $En^k$ is finite. Hence $\lim\,(1 - P_N)Q(N) = 0$. Because of (1) the integrals in the second term of (5) are bounded as $N \to \infty$. Now set $s = 0$ in (5) and then let $N \to \infty$. Since $\phi(0) = 1$, the second term of (5) approaches 0 and the limit of the first term is just the left side of (3).

For the case of a Wald sequential process, Stein [4] has shown that all moments of $n$ are finite. In this case (3) holds whenever $Ez^k$ is finite.

## REFERENCES

[1] DAVID BLACKWELL AND M. A. GIRSHICK, "On functions of sequences of independent chance vectors, with applications to the problem of the random walk in $k$ dimensions," *Annals of Math. Stat.*, Vol. 17 (1946), p. 310.

[2] ABRAHAM WALD, "On cumulative sums of random variables," *Annals of Math. Stat.*, Vol. 15 (1944), p. 283.

[3] ABRAHAM WALD, "Differentiation under the expectation sign in the fundamental identity of sequential analysis," *Annals of Math. Stat.*, Vol. 17 (1946), p. 493.

[4] CHARLES STEIN, "A note on cumulative sums," *Annals of Math. Stat.*, Vol. 17 (1946), p. 498.

# A UNIQUENESS THEOREM FOR UNBIASED SEQUENTIAL BINOMIAL ESTIMATION

BY L. J. SAVAGE[1]

*University of Chicago*

In a recent note [1], J. Wolfowitz extended some of the results of a paper by Girshick, Mosteller and Savage [2] on sequential binomial estimation. The present note carries one of Wolfowitz's ideas somewhat further. The nomenclature of [1] and [2] will be used freely. The concept of "doubly simple region" introduced in [1] and assumed there only in the hypothesis of Theorem 3, will here be shown to be unnecessarily restrictive. In so doing, we find that sim-

[1] The author is a Rockefeller fellow at the Institute of Radiobiology and Biophysics, University of Chicago.

plicity is not only a necessary (cf. Theorem 4 of [2]) but also a sufficient condition that $\hat{p}$ be the unique unbiased estimate of $p$ for a closed region.

LEMMA. *If $R$ is simple there is at most one bounded unbiased estimate of any given function of $p$.*

PROOF. If the lemma were false, there would be a non-trivial bounded unbiased estimate of zero, i.e., $m(\alpha)$ such that $| m(\alpha) |$ is bounded by a constant $m^*$, $m(\alpha)$ not identically zero and $E(m(\alpha) \mid p) \equiv 0$.

$$E(m(\alpha) \mid p) = \sum m(\alpha)k(\alpha)p^y q^x = 0. \tag{1}$$

and $m(\alpha)$ not identically zero. Since $R$ is simple we may assume (much as in the proof of Theorem 6 of [2]) that we have a boundary point such that $m(\alpha_0) \neq 0$, $\alpha_0$ is below all accessible points of its own index and also below every other $\alpha$ for which $m(\alpha) \neq 0$. Therefore

$$| m(\alpha_0) | \, k(\alpha_0)p^{y_0} q^{x_0} = \Big| \sum_{y > y_0} m(\alpha)k(\alpha)p^y q^x \Big| \leq m^* \sum_{y > y_0} k(\alpha)p^y q^x. \tag{2}$$

Let $M$ denote the set of all accessible points and boundary points at which $x < x_0$ and $y = y_0 + 1$. There are at most $x_0$ points in $M$, say $\beta_1, \cdots, \beta_n$. Considering the way in which $\alpha_0$ has been chosen, every path from (0, 0) to an $\alpha$ for which $y > y_0$ passes through or to at least one point of $M$. Therefore when $y > y_0$

$$\begin{aligned} P(\alpha) = k(\alpha)p^y q^x &= P(\alpha \mid M)P(M) \\ &\leq P(\alpha \mid M) \sum_1^n k(\beta_j)p^{y_0+1} q^{x_j} \\ &\leq p^{y_0+1} \sum_1^n k(\beta_j)P(\alpha \mid M). \end{aligned} \tag{3}$$

From inequalities (2) and (3).

$$\begin{aligned} | m(\alpha_0) | \, k(\alpha_0)p^{y_0} q^{x_0} &\leq m^* p^{y_0+1} \left\{ \sum_1^n k(\beta_j) \right\} \sum_{y > y_0} P(\alpha \mid M) \\ &\leq m^* p^{y_0+1} \sum_1^n k(\beta_j). \end{aligned} \tag{4}$$

But it is impossible that (4) should be satisfied for small $p$.

Combining the Lemma with Theorem 4 of [2] we have the

THEOREM. *A necessary and sufficient condition that $\hat{p}(\alpha)$ be the unique proper (bounded) and unbiased estimate of p for a closed region R is that R be simple.*

The sufficiency part of this Theorem extends Theorem 3 of [1] from doubly simple regions to simple regions.

The author is indebted to J. Wolfowitz for his valuable suggestions in connection with the present note.

REFERENCES

[1] J. Wolfowitz, "On sequential binomial estimation," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 489-493.

[2] M. A. Girshick, Frederick Mosteller, and L. J. Savage, "Unbiased estimates for certain binomial sampling problems with applications." *Annals of Math. Stat.*, Vol. 17 (1946), pp. 13-23.

# ACKNOWLEDGEMENT OF PRIORITY

By H. E. Robbins

*University of North Carolina*

At the time of publication of my papers on the measure of a random set (*Annals of Math. Stat.*, Vol. 15 (1944), pp. 70–74; Vol. 16 (1945), pp. 342–347), I was unaware that the theorem on page 72 of the first paper, which affords a means of computing the expected value of the measure, had already been found by A. Kolmogoroff. (*Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, Berlin, 1933, p. 41). I wish to take this opportunity of acknowledging Kolmogoroff's priority, which was pointed out by Prof. Henry Scheffé.

# ABSTRACTS OF PAPERS

Presented on January 25, 1947, at the Atlantic City meeting of the Institute

## 1. A Test of Significance of the Coefficient of Rank Correlation for more than Thirty Ranked Items. NILAN NORRIS, Hunter College.

Hotelling and Pabst (*Annals of Math. Stat.*, Vol. 7 (1936), p. 37) have suggested the use of the Tchebycheff inequality as an approximation for testing the significance of the coefficient of rank correlation in cases where the number of ranked items is too large to enable exact probabilities to be computed directly. A table prepared in accordance with this suggestion indicates that for values of the coefficient of rank correlation larger than .50 there is a wide range of corresponding numbers of ranked items greater than thirty for which at least the five per cent level of significance is satisfied.

For certain types of applications the conservativeness of the Tchebycheff test may be a virtue rather than a limitation.

## 2. A Generalized T Measure of Multivariate Dispersion. HAROLD HOTELLING, University of North Carolina.

The problem of combining errors in two or more dimensions to measure the accuracy of firing and bombing is similar to problems occurring in industrial quality control where different measures of quality are applied to the same article, and to problems in mental testing and other fields. If the covariances were known a priori, the solution optimum in certain senses, for a multivariate normal distribution, would be the use of $\chi^2 = \Sigma\Sigma\lambda_{ij}x_ix_j$, where $[\lambda_{ij}]^{-1}$ is the covariance matrix and $x_i$ is the deviation in the $i$th dimension. Since the covariances must in all known practical cases be estimated from a preliminary sample with (say) $n$ degrees of freedom, $\chi^2$ may be replaced by $T^2 = \Sigma\Sigma l_{ij}x_ix_j$, where $[l_{ij}]^{-1}$ is the estimated covariance matrix. This is the same $T$ introduced by the author in 1931 as a generalization of the Student ratio $t$, and has the same distribution. Upon adding together the values of $T^2$ for different cases (e.g. for different bombs dropped with the same bombsight), a combined measure $T_0^2$ of over-all excellence (e.g. of the bombsight), is obtained. $T_0^2$ like $\chi^2$, can be broken down into components meaningful with respect to the causal system, specifically in relation to possible sources of excessive discrepancy. Thus, if $\bar{x}_i$ is the $i$h coordinate of the centroid, or mean point of impact, of $m$ bombs, we may write $T_M^2 = \Sigma\Sigma l_{ij}\bar{x}_i\bar{x}_j$, $T_D^2 = T_0^2 - T_M^2$. Then $T_D$ is a function only of deviations from the mean point of impact. Asymptotically (for large $n$), $T_0$, $T_M$ and $T_D$ have the $\chi$ distribution with $m$, 2 and $m - 2$ degrees of freedom respectively. But the untrustworthiness of the $\chi$ distribution as an approximation is evident even with $n$ as large as 256, for which case calculations have been made. The exact distributions of $T_0$ and $T_D$ are ascertained when the number of variates $p$ is 2, and the probability integrals are expressed as linear functions of two incomplete beta functions. In fact, $T_0^2/M$ equals the sum of the roots of a determinantal equation of the form $| A - \lambda\beta | = 0$, where $A$ and $B$ are sample covariance matrices with $n$ and $m$ degrees of freedom respectively, and a similar relation holds for $T_D^2$ with $m$ replaced by $m - 2$. $T_0$ and $T_M$ have the distribution published in 1931, with probability integral expressible in terms of a single incomplete beta function or the variance ratio distribution. It is shown that such parameters as the circular mean deviation are best estimated with the help of the $T$ measures, not directly by averaging individual circular deviations.

## 3. Asymptotic Properties of Maximum and Quasi-Maximum Likelihood Estimates. HERMAN RUBIN, Cowles Commission for Research in Economics.

The results of J. L. Doob (*Trans. Am. Math. Soc.*, Vol. 36 (1934), pp. 759–775) on consistency of maximum likelihood estimates, are generalized and extended to arbitrary measure spaces. In some special cases, results on asymptotic normality of maximum likelihood estimates can be generalized to quasi-maximum likelihood estimates (estimates based on the assumption of a likelihood function which need not be the true function).

4. **The Asymptotic Distribution of the Range.** E. J. GUMBEL, Newark College of Engineering.

The asymptotic distribution of the range $w$ for initial unlimited distributions of the exponential type is obtained by convolution of the asymptotic distributions of the two extremes. Let $\alpha$ and $u$ be the parameters of the distributions of the extremes for a symmetrical variate, and let $R = \alpha(w - 2u)$ be the reduced range. Then the probability $\Psi(R)$ of the reduced range is subject to the differential equation $\Psi'' + \Psi' - \Psi \exp(-R) = 0$ which may be transformed into Bessel's equation of the first order by the substitutions $R = 2(\log 2 - \log z)$, and $\Psi = zU$. The solution is $\Psi(R) = zK_1(z)$ for the asymptotic probability, and $\psi(R) = (z^2/2)K_0(z)$ for the asymptotic distribution, $K_0(z)$ and $K_1(z)$ being the modified Bessel function of the second kind of orders zero and unity. Thus tables of $\Psi(R)$ and $\psi(R)$ may be calculated for any symmetrical distribuion of the exponential type. The distribution of the range $w$ for normal samples of size 10 is already very close to the asymptotic distribution provided that the parameters $\alpha$ and $u$ are determined from the mean and the standard deviation of the range. This method permits the calculation of the distribution of the range for normal samples of any size larger than 10.

5. **The Corner Test for Association.** JOHN W. TUKEY, Princeton University, and PAUL S. OLMSTEAD, Bell Telephone Laboratories.

*Construction.* In a scatter diagram, draw the two medians, that is, the median of the $x$ values without regard to the values of $y$, and the median of the $y$ values without regard to the values of $x$. Think of the four quadrants thus formed as being labelled $+$, $-$, $+$, $-$ in order, so that the two positive quadrants lie along one diagonal and the two negative along the other. Beginning at the right-hand side of the diagram, count in along the observations until forced to cross the horizontal median. Write down the number of observations met before this crossing, attaching the sign, $+$, if they lay in the $+$ quadrant, and the sign, $-$, if they lay in the $-$ quadrant. Repeat this process, moving up from below, moving to the right from the left, and moving down from above. The quantity to be used in the test is the algebraic sum of the four numbers thus written down.

*Distribution.* The exact distribution of this quantity when no association is present and no two $x$'s and no two $y$'s are alike is almost independent of sample size over the range of values where it is apt to be used. For example, a sum of 9 or more is expected less than one time in ten for all samples of size 6 or more; a sum of 15 or more, less than one time in 100 for all samples of size 10 or more; and a sum of 21 or more, less than one time in 1000 for all samples of size 14 or more. Even for infinite sample size, the sums for these fractions become only 9, 14, and 19, respectively.

*Extensions.* The same ideas that underlie the outside corner test for two variables may be extended in several ways to give tests for various types of association among three or more variables.

6. **Consistent Estimates Based on Partially Consistent Observations, with Particular Reference to Structural Relations.** J. NEYMAN AND ELIZABETH L. SCOTT, University of California.

Let $\{X_n\}$ be a sequence of independent random variables and let $F_i$ denote the distribution of $X_i$. Each distribution $F_i$ is assumed to depend on unknown parameters. If a parameter $\theta$ appears in an infinity of distributions $F_i$, it is called *structural*. Otherwise, it is *incidental*. The sequence $\{X_n\}$ is called *consistent* if $\{F_n\}$ has no incidental parameters. $\{X_n\}$ is called *partially consistent* if $\{F_n\}$ has both structural and incidental parameters.—Problem of fitting a straight line when both variables are subject to errors is that of a partially consistent series of observations. Let $\xi$ and $\eta = \alpha + \beta\xi$ be two linearly connected quantities, perhaps related to particular stars, where $\alpha$ and $\beta$ are unknown. The values $\xi_i$ and $\eta_i$ corresponding to the $i$th star, $(i - 1, 2, \cdots, s)$, are unknown. The observations provide measurements $x_{ij}$ of $\xi_i$, $(j - 1, 2, \cdots, m_i)$, and measurements $y_{ik}$, $(k = 1, 2, \cdots, n_i)$, of $\eta_i$. Both $m_i$ and $n_i$ are bounded and small. On the other hand, $s$ may be considered as increasing without limit.—Assume that the $x_{ij}$ and the $y_{ik}$ are normally distributed with variances $\sigma_1^2$ and $\sigma_2^2$ and means $\xi_i$ and $\eta_i$ respectively. Then the totality of observations will form a partially consistent system with the structural parameters $\alpha$, $\beta$, $\sigma_1$ and $\sigma_2$ and with $\xi_i$ as incidental parameters.—If the observable random variables are only partially consistent, then the maximum likelihood estimates of the structural parameters (a) need not be consistent, (b) even if they are consistent and asymptotically normal, alternative estimates may exist which have the same properties but smaller asymptotic variances.—Consistent estimates of structural parameters may be obtained from "modified" equations of maximum likelihood. The lower bound of the variance of estimates of structural parameters, provided by the Cramér-Rao inequality, is attained only on certain conditions which are both necessary and sufficient.

# NEWS AND NOTICES

*Readers are invited to submit to the Secretary of the Institute news items of interest*

## Personal Items

Dr. Paul H. Anderson has been appointed Economic Analyst with the Marketing Division, Office of Domestic Commerce, Department of Commerce, Washington.

Dr. Gilbert W. Beebe is now with the Division of Medical Sciences, National Research Council, Washington.

Professor Harald Cramér, Director of the Institute of Mathematical Statistics of the University of Stockholm, was awarded the degree of Doctor of Science, *honoris causa*, by Princeton University on February 22, 1947. Professor Cramér has acted as Visiting Professor of Mathematics at Princeton University and Yale University during the academic year 1946–'47. He will be at the University of California at Berkeley during the 1947 Summer Session.

Dr. Paul M. Densen has accepted a position with the Division of Medical Research Statistics, Bureau of Medicine and Surgery, Veterans Administration, Washington.

Mr. M. V. Divatia is now in charge of the office of the Statistician and Economic Adviser and Under-Secretary to the Government of Sind, Karachi, India.

Mr. Clarence B. Fine, formerly with the Office of Price Administration, has transferred to the Bureau of Old-Age and Survivors Insurance, Social Security Administration, where he is employed as a Sampling Expert.

Prof. Charles C. Grove was appointed Visiting Lecturer in Mathematics at the University of Pennsylvania for the spring semester.

Assoc. Prof. E. E. Haskins of Northeastern University has been appointed to an assistant professorship at the Army Air Forces Institute of Technology, Wright Field, Dayton, Ohio.

Prof. Roger Lessard of the Hull Technical School has accepted a position at the Ecole Polytechnique, Montreal.

Mr. Edward D. Lowery is now a member of the Research Department, Winchester Arms Company, New Haven, Connecticut.

Professor H. B. Mann of Ohio State University has been awarded the Frank Nelson Cole prize in the Theory of Numbers for 1946.

Dr. Margaret P. Martin has been appointed to an assistant professorship in the Department of Preventive Medicine and Public Health, Vanderbilt University Medical School, Nashville, Tennessee.

Dr. A. L. O'Toole is at present employed by the Veterans Administration in the Washington headquarters, as Acting Chief of the Administrative Analysis Division in the Research Service. Dr. O'Toole was released from the Navy on September 23, 1946, to inactive duty in the U. S. Naval Reserve, with the rank of Commander. Dr. O'Toole served for nearly four years in the Navy, in important administrative and statistical work for the Commander South Pacific Area and South Pacific Force. He will be remembered as having been with Admiral Halsey's Pacific Fleet, and was awarded the Bronze Star Medal. At the time of his release, he was Chief Staff Officer for Commander South Pacific Area and South Pacific Force.

Mr. I. B. Perrott, since his demobilization from the British Army, has been Lecturer in Mathematics at the College of Technology and Commerce, Leicester, England.

Mr. J. S. Ripandelli is now with the Actuarial Department of the Jefferson Standard Life Insurance Company of Greensboro, North Carolina.

Dr. Ronald W. Shephard of the University of California has been appointed to the staff of the Department of Mathematics, New York University.

Mr. John R. Stehn is now a member of the Research Laboratory of the General Electric Company, Schenectady, New York.

Dr. Charles W. Vickery, formerly of Ohio State University, is engaged in work as a Research Consultant in New York City.

Miss Margaret Jeannin Dix, of the University of California Statistical Laboratory, died an accidental death at her home in Berkeley on June 20, 1946.

Mr. Albert M. Freeman, of the Boston Fiduciary and Research Association, died May 20, 1946.

Dr. Walter Schilling, of the Stanford University Hospital, died suddenly in San Francisco, December 16, 1946.

## Summer Statistical Session at the University of California at Berkeley

The important advances in the theory of statistics during the war and especially the unprecedented growth in the fields of application have created a strong demand for trained statisticians to fill both the research and the teaching positions all over the country. Since in many cases the war time education had to be somewhat sketchy, unsystematic, and not very conducive to a thorough coverage of the vast material, it is felt that a relatively brief set of courses on a rather advanced level would be beneficial to many persons, both those who already hold research or teaching positions in statistics, as well as those who prepare for higher degrees.

With this object in mind, the University of California at Berkeley is offering a set of statistical courses during the Summer Session, June 23rd to August 2nd, 1947. There will be three courses: (i) General Theory of Random Variables and Frequency Distributions, by Harald Cramér of the University of Stockholm;

(ii) Problems of Testing Hypotheses and of Estimation, by J. Neyman, University of California, Berkeley; and (iii) Seminar Course. The last will be given by seven scholars, each giving two hours of lectures, as follows:

| | |
|---|---|
| 1. Statistical Astronomy. | R. J. Trumpler |
| 2. Orthogonal Polynomials and Problems of Moments. | G. Szegö |
| 3. Methods of Calculation. | V. F. Lenzen |
| (a) Gibbs' Methods in Statistical Mechanics. | |
| (b) Darwin-Fowler Method of Statistics. | |
| 4. Large Scale Sampling Surveys. | P. C. Mahalanobis |
| 5. Statistical Problems Arising in Nuclear Physics Measurements. | R. Serber |
| 6. Problems of Population Genetics. | S. Emerson |
| 7. Interactions between Industrial Problems and Mathematical Statistics. | H. Scheffé |

The purpose of the Seminar Course is to introduce the students either to branches of pure mathematics contingent on mathematical statistics but not ordinarily taught in the universities or to various fields of knowledge offering fruitful fields for statistical studies.

## Summer Statistical Session at Virginia Polytechnic Institute

A Summer Statistical Session will be held at Virginia Polytechnic Institute, Blacksburg, Virginia, August 5 to September 5, 1947. This Session will be sponsored jointly by Virginia Polytechnic Institute, University of North Carolina, University of Michigan, Iowa State College, and the Federal Bureau of Agricultural Economics.

The faculty will consist of; Walter A. Hendricks, B.A.E., U.S.D.A.; Renis Likert, University of Michigan; H. L. Lucas, University of North Carolina; Maurice G. Kendall, England; George W. Snedecor, Iowa State College; Frank Yates, Rothamsted Experiment Station, England; Earl E. Houseman, B.A.E., U.S.D.A.; Raymond J. Jessen, Iowa State College, and Boyd Harshbarger, Virginia Polytechnic Institute.

The following courses will be offered for credit: Engineering Statistics; Statistical Methods; Design of Animal Experiments; Schedule Design and Interview Techniques for Sample Surveys; Sampling Design and Analysis; Mathematical Theory of Sampling; Seminar; Mathematical Statistics, and Experimental Design.

In addition to the faculty, probable Seminar speakers are: W. F. Callendar, W. G. Cochran, Miss Gertrude M. Cox, W. E. Deming, George Gallup, M. H. Hansen, Harold Hotelling, Arnold King, and Charles F. Sarle.

Inquiries regarding the Summer Session should be addressed to Boyd Harshbarger, Professor of Statistics, Summer Statistical Session, Virginia Polytechnic Institute, Blacksburg, Virginia.

## New Members

*The following persons have been elected to membership in the Institute (January 1 to February 28, 1947):*

**Asofsky, Samuel,** B.S. (C.C.N.Y.) Stat., National Jewish Welfare Board, *1256 E. 13 St., Brooklyn 30, N. Y.*

**Auer, Richard M.,** A.M. (Columbia) Instr. in Math., State Teachers Coll., Montclair, N. J., *88 No. 16 St., East Orange*

**Bakan, David,** M.A. (Indiana) Chief Stat., Comm. on Selection and Training of Aircraft Pilots, National Research Council, *259 Natatorium, Ohio State Univ., Columbus 10, Ohio*

**Beatty, Glenn H.,** A.B. (Ohio State) Grad. student and Fellow, Iowa State College, *Station A, General Delivery, Ames, Iowa*

**Campbell, Wallace A.,** B.S. (Columbia) Stat .Analyst, War Assets Administration, *483 Washington Ave., Brooklyn 16, N. Y.*

**Cella, Francis R.,** M.A. (Kentucky) Assoc. Prof. of Statistics and Director, Bur. of Business Research, Univ. of Oklahoma, Norman, Okla.

**Chapman, Douglas G.,** M.A. (Toronto) Asst. Prof. of Math., Univ. of British Columbia, Vancouver, Canada

**Cheydleur, Benjamin F.,** B.A. (Wisconsin) Chief, Mechanized Analysis, Naval Ordnance Lab., *602 Avenue E, District Heights, Washington 19, D. C.*

**Coombs, Clyde H.,** Ph.D. (Chicago) Ass't Prof. of Psychology, and Research Psychologist, Institute for Human Adjustment, Univ. of Michigan, Ann Arbor, Mich., *1027 E. Huron*

**Corton, Edward L., Jr.,** M.B.A. (Chicago) Grad. student, Iowa State Coll., *803 Hodge Ave., Ames, Iowa*

**Davis, Harold.,** A.B. (Brooklyn Coll.) Stat., Navy Dept., *416—33 St., S.E., Washington, D. C.*

**Dutton, Arthur M.,** B.S.E.E. (Iowa State) Grad. Fellow, Mathematics Dept., Iowa State Coll., Ames, Iowa

**Fay, Edward A.,** A.M. (Harvard) Grad. student, Univ. of California, Berkeley, *415 South 17th St., Apt. 2B, Richmond, Calif.*

**Flanagan, John C.,** Ph.D. (Harvard) Prof. of Psychology, Univ. of Pittsburgh, Pittsburgh 13, Pa.

**Gardner, Eric F.,** Ed.M. (Boston Teachers) Teaching Fellow and Milton Fellow, Grad. School of Educ., Harvard Univ., Cambridge, Mass., *Walker House, 40 Quincy St.*

**Gerende, Lincoln J.,** C.Ph.M., U. S. Navy, *Naval Medical Res. Institute, National Naval Medical Center, Bethesda 14, Md.*

**Grossman, Evelyn,** M.A. (Columbia) Stat., U. S. Dept. of Agriculture, *6401—14 St., N. W., Washington 12, D. C.*

**Hill, Edwin A., Jr.,** M.A. (Columbia) Instr. in Math., Coll. of the City of N. Y., *50 West 67 St., New York 23, N. Y.*

**Horton, H. Burke,** M.B.A. (Texas) Senior Transport Analyst, *2906 Naylor Rd., S. E., Washington 20, D. C.*

**Horvitz, Daniel G.,** B.S. (Mass. State) Grad. student, Iowa State Coll., *2137 Country Club Blvd., Ames, Iowa*

**Ikhtiar-ul-Mulk, S. M.,** M.A. (Punjab, India) Grad. student, Princeton Univ., *Graduate College, Princeton, N. J.*

**Jaeger, Carol M.,** B.A. (Dubuque) Statistician, *1300 Columbia Terrace, Peoria 5, Ill.*

**Jessen, Raymond J.,** Ph.D. (Iowa State) Res. Assoc. Prof., Iowa State College, and Agric. Statistician, U.S.D.A., *Statistical Lab., Iowa State Coll., Ames, Iowa*

**Kinzer, Mrs. Lydia Greene,** M.A. (Kansas) Ass't Instr. in Math., Ohio State Univ., *585 East Town Street, Columbus 15, Ohio*

**Langenhop, Carl E.,** M.S. (Iowa State) Instr. in Math., Iowa State Coll., *Apt. 3, Cranford Annex, Ames, Iowa*

**Lowy, Melitta E.,** A.B. (Hunter) Statistician, Grad. student, Columbia Univ., *645 West End Ave., New York 25, N. Y.*

**Mattila, Sakari,** Fil.Mag. (Helsinki) High School of Commerce, Helsinki, Finland

**Mayerson, Allen L.,** B.S. (Michigan) Grad. student and Teaching Fellow, Univ. of Mich., *1302 Packard St., Ann Arbor, Mich.*

**McCreary, Garnet E.,** M.A. (Queen's Univ.) Research Fellow, Statistical Lab., Iowa State Coll., Ames, Iowa

**McMillan, Olan T.,** M.A. (Michigan) Instr. in Math., Michigan State Coll., East Lansing, Mich.

**Morris, Edward B.,** A.B. (Indiana) Statistician, U. S. Bur. of Labor Statistics, *1915 Ridge Place S. E., Washington 20, D. C.*

**Moshman, Jack,** B.A. (New York) Tutor in Math., Queens Coll., Flushing, N. Y., *125–09 Liberty Ave., Richmond Hill 19*

**Natrella, Mrs. Mary G.,** B.A. (Pennsylvania) Statistician, Bureau of Ships, Navy Dept., *1210—12th St., N. W. Washington 5, D. C.*

**Neal, T. Ellison,** A.B. (Geo. Washington) Statistician, Textile Dev. Dept., U. S. Rubber Co., Hogansville, Ga.

**Noble, Carl E.,** Ph.D. (Iowa) Quality Methods Engineer, Kimberly Clark Corp., Lakeview Mill, Neenah, Wis.

**Ostle, Bernard,** M.A. (British Columbia) Teaching Ass't, School of Bus. Adm., Univ. of Minnesota, Minneapolis, Minn.

**Oxtoby, Toby E.,** B.A. (Iowa) Grad. Ass't, Dept. of Psychology, State Univ. of Iowa, Iowa City, Iowa

**Peisakoff, Melvin P.,** Student, Princeton Univ., *34 North West College, Princeton, N. J.*

**Rothschild, Colette,** (Ecole Normale Superieure) Attachee de Recherches au Centre National de la Recherche Scientifique, *43 rue Madame, Paris VI*ᵉ*, France*

**Slonim, Morris J.,** M.B.A. (Harvard) Statistician, Bureau of Labor Statistics, *210 Wayne Place S. E., Washington 20, D. C.*

**Soler, Reuben I.,** B.B.A. (C.C.N.Y.) Statistician, Food and Drug Administration, *246 Portland St., S. E., Washington, D. C.*

**Stouffer, Samuel A.,** Ph.D. (Chicago) Prof. of Sociology and Director of the Laboratory of Social Relations, Emerson Hall, Harvard Univ., Cambridge, Mass.

**Teicher, Henry,** B.A. (Iowa) Graduate student, Columbia Univ., *139 Osborne Terrace, Newark, N. J.*

**Tiedeman, David V.,** M.A. (Rochester) Instr. in Educ., Grad. School of Educ., Harvard Univ., *Walker House, 40 Quincy St., Cambridge 38, Mass.*

**Tintner, Gerhard,** Ph.D. (Vienna) Prof. of Economics and Mathematics, Iowa State Coll., Ames, Iowa

**Weiss, Eleanor S.,** Ed.M. (Boston Teachers) Teaching Fellow, Grad. School of Educ., Harvard Univ., *2005 Commonwealth Ave., Brighton 35, Mass.*

**Wilson, William A., Jr.,** A.B. (California) Teaching Ass't in Psychology, Univ. of Calif., Berkeley 4, Calif.

**Woodell, Allan D.,** A.B. (N. Y. State Teachers, Albany) Graduate student in math., Univ. of Mich., *425 Church St., Ann Arbor, Mich.*

*Omitted from 1946 lists of new members:*

**Feraud, Prof. Lucien,** Faculte des Sciences Economiques et Sociales, Univ. de Geneve, *24 rue Henri Mussard, Genève, Switzerland*

## REPORT ON THE ATLANTIC CITY MEETING OF THE INSTITUTE

The Ninth Annual Meeting of the Institute of Mathematical Statistics was held at Atlantic City, New Jersey, on Friday and Saturday, January 24–25, 1947. The meeting was held in conjunction with meetings of the American Economic Association, American Statistical Association, and the Econometric Society. The following 154 members of the Institute attended the meeting:

Beatrice Aitchison, F. L. Alt, R. L. Anderson, T. W. Anderson, K. J. Arrow, Max Astrachan, B. M. Bennett, Joseph Berkson, A. J. Berman, C. I. Bliss, Paul Boschan, A. E. Brandt, M. F. Bresnahan, Philip Brown, O. P. Bruno, R. W. Burgess, O. K. Buros, B. H. Camp, F. R. Cella, Uttam Chand, K. L. Chung, C. W. Churchman, P. C. Clifford, W. J. Cobb, W. G. Cochran, F. G. Cornell, D. R. Cowan, Harald Cramèr, J. H. Curtiss, J. F. Daly, G. B. Dantzig, D. G. Deihl, D. B. DeLury, B. W. Dempsey, H. F. Dorn, F. W. Dresch, A. J. Duncan, David Durand, P. S. Dwyer, Churchill Eisenhart, W. D. Evans, Will Feller, C. D. Ferris, Irving Fisher, L. R. Frankel, M. A. Geisler, Leon Gilford, M. A. Girshick, C. H. Graves, K. E. Greene, S. W. Greenhouse, F. E. Grubbs, E. T. Gumbel, Margaret Gurney, Louis Guttman, Trygve Haavelmo, K. W. Halbert, M. H. Hansen, Miriam S. Harold, T. E. Harris, Boyd Harshbarger, Bernard Hecht, Wassily Hoeffding, H. B. Horton, Harold Hotelling, E. E. Houseman, Helen M. Humes, Leonid Hurwicz, Seymour Jablon, R. W. James, R. J. Jessen, H. L. Jones, Alice S. Kaitz, H. B. Kaitz, L. S. Kellogg, H. S. Konijn, Tjalling Koopmans, C. F. Kossack, R. L. Kozelka, D. H. Leavens, Howard Levene, J. E. Lieberman, Rensis Likert, S. B. Littauer, Irving Lorge, P. J. McCarthy, P. W. McGann, F. E. McIntyre, H. F. MacNeish, J. D. Maddrill, Jacob Marschak, Max Millikan, A. M. Mood, Mrs. Margaret Moore, J. W. Morse, J. E. Morton, Frederick Mosteller, D. N. Nanda, P. M. Neurath, Jerzy Neyman, M. L. Norden, Nilan Norris, H. W. Norton, P. S. Olmstead, E. G. Olds, Sophie Rakesky, Chester Rapkin, Olav Reiersol, W. A. Reynolds, P. R. Rider, C. F. Roos, A. C. Rosander, Ernest Rubin, Herman Rubin, P. J. Rulon, Frank Saidel, Marion M. Sandomire, Max Sasuly, F. E. Satterthwaite, E. D. Schell, E. M. Schrock, D. H. Schwartz, G. R. Seth, L. W. Shaw, W. A. Shewhart, J. H. Smith, R. T. Smith, Leslie E. Simon, Milton Sobel, C. M. Stein, G. T. Steinberg, Joseph Steinberg, H. W. Steinhaus, F. F. Stephan, A. P. Stergion, M. S. Stevens, G. J. Stigler, S. A. Stouffer, Zenon Szatrowski, B. J. Tepping, J. W. Tukey, D. F. Votaw, Jr., Helen M. Walker, J. H. Watkins, Louis Weiner, Samuel Weiss, S. S. Wilks, Elizabeth W. Wilson, C. P. Winsor, J. Wolfowitz, M. A. Woodbury, Holbrook Working, C. A. Wright, and T. O. Yntema.

The first session, a joint session with the Econometric Society and the Biometrics Section of the American Statistical Association, was held at two o'clock on Friday afternoon, and was devoted to the topic, *Applications of Statistical Techniques to Agricultural Economics.* Holbrook Working of Stanford University presided. The following four papers were presented:

1. *Use of Variance Components in the Analysis of Market Differentials in Hog Prices.* R. L. Anderson, University of North Carolina.
2. *An Application of the Analysis of Variance in the Economic Evaluation of Production.* Boyd Harshbarger, Virginia Polytechnic Institute.
3. *A Model of the Economic Interdependence between Agriculture and the National Economy.* Trygve Haavelmo, Cowles Commission for Research in Economics.
4. *The Reduced-Form Method for Estimating Simultaneous Economic Relationships.* M. A. Girschick, Bureau of the Census.

The session concluded with a discussion of these papers by T. W. Anderson, Columbia University; Milton Friedman, University of Chicago; and, Harold Hotelling, University of North Carolina.

At 8 o'clock on Friday evening there was a joint session with the Econometric Society and the American Statistical Association, on the topic, *When is the Analysis of Variance Useful in Economic Research?* Arthur R. Tebbutt of Northwestern University presided, and the following three papers were presented:

1. *The Advantages of the Analysis of Variance for Research and Managerial Control Purposes.* Harry Pelle Hartkemeier, University of Missouri.
2. *Estimation of Economic Relationships and Multivariate Regression.* Leonid Hurwicz, Iowa State College.
3. *Nonstandard Forms of Variance Analysis.* W. Allen Wallis, University of Chicago.

There was discussion of these papers by Tjalling Koopmans, Cowles Commission for Research in Economics: Gerhard Tintner, Iowa State College; and, J. W. Tukey, Princeton University.

At 10 o'clock on Saturday morning there was a joint session with the American Statistical Association devoted to the topic, *Use of Ordered Observations in Statistical Analysis*, with Harold Hotelling of the University of North Carolina as chairman. The following two papers were presented:

1. *Estimation of Parameters by Use of Order Statistics.* Frederick Mosteller, Harvard University.
2. *Tolerance Limits.* Jacob Wolfowitz, Columbia University.

There was discussion of these papers by John H. Smith, Bureau of Labor Statistics; Howard L. Jones, Illinois Bell Telephone Company; and J. W. Tukey, Princeton University.

At the Saturday morning session one contributed paper of the Institute of Mathematical Statistics was also presented, by E. J. Gumbel, Newark College of Engineering, on the topic: *The Asymptotic Distribution of the Range.*

The Institute's session at 2 o'clock Saturday afternoon was devoted to contributed papers. W. G. Cochran, president of the Institute, presided, and the following four papers were presented:

1. *A Test of Significance of the Coefficient of Rank Correlation for More than Thirty Ranked Items.* Nilan Norris, Hunter College.
2. *A Generalized T Measure of Multivariate Dispersion.* Harold Hotelling, University of North Carolina.
3. *Asymptotic Properties of Maximum and Quasi-Maximum Likelihood Estimates.* Herman Rubin, Cowles Commission for Research in Economics.
4. *The Corner Test for Association.* J. W. Tukey, Princeton University, and Paul Olmstead, Bell Telephone Laboratories.

Abstracts of these papers appear elsewhere in this issue.

Following the session on contributed papers, Professor Jerzy Neyman of the University of California gave an invited address on the topic: *On Consistent Estimates, with Particular Reference to Structural Relations between Several Variables all Subject to Random Error*. A discussion of this address followed, by Miss E. L. Scott, University of California; A. Wald, Columbia University; and Tjalling Koopmans, Cowles Commission for Research in Economics.

The meeting closed with the annual business meeting of the Institute, which was held at 5 p.m. on Saturday in Haddon Hall. Reports by the President, Secretary-Treasurer, and Editor were followed by the election of officers for 1947: Will Feller, President; Morris H. Hansen and John H. Curtiss, Vice-Presidents; and Paul S. Dwyer, Secretary-Treasurer.

P. S. DWYER,
*Secretary.*